ASYMPTOTICS OF MARKOV KERNELS AND THE TAIL CHAIN

SIDNEY I. RESNICK AND DAVID ZEBER

ABSTRACT. An asymptotic model for extreme behavior of certain Markov chains is the "tail chain". Generally taking the form of a multiplicative random walk, it is useful in deriving extremal characteristics such as point process limits. We place this model in a more general context, formulated in terms of extreme value theory for transition kernels, and extend it by formalizing the distinction between extreme and non-extreme states. We make the link between the update function and transition kernel forms considered in previous work, and we show that the tail chain model leads to a multivariate regular variation property of the finite-dimensional distributions under assumptions on the marginal tails alone.

1. INTRODUCTION

A method of approximating the extremal behavior of discrete-time Markov chains is to use an asymptotic process called the *tail chain* under an asymptotic assumption on the transition kernel of the chain. Loosely speaking, if the distribution of the next state converges under some normalization as the current state becomes extreme, then the Markov chain behaves approximately as a multiplicative random walk upon leaving a large initial state. This approach leads to intuitive extremal models in such cases as autoregressive processes with random coefficients, which include a class of ARCH models. The focus on Markov kernels was introduced by Smith [24]. Perfekt [18, 19] extended the approach to higher dimensions, and Segers [23] rephrased the conditions in terms of update functions.

Though not restrictive in practice, the previous approach tends to mask aspects of the processes' extremal behaviour. Markov chains which admit the tail chain approximation fall into one of two categories. Starting from an extreme state, the chain either remains extreme over any finite time horizon, or will drop to a "non-extreme" state of lower order after a finite amount of time. The latter case is problematic in that the tail chain model is not sensitive to possible subsequent jumps from a non-extreme state to an extreme one. Previous developments handle this by ruling out the class of processes exhibiting this behaviour via a technical condition, which we refer to as the *regularity condition*. Also, most previous work has assumed stationarity, since interest focused on computing the extremal index or deriving limits for the exceedance point processes, drawing on the theory established for stationary processes with mixing by Leadbetter et al. [17]. However, stationarity is not fundamental in determining the extremal behaviour of the finite-dimensional distributions.

We place the tail chain approximation in the context of an extreme value theory for Markovian transition kernels, which *a priori* does not necessitate any such restrictions on the class of processes to which it may be applied. In particular, we introduce the concept of boundary distribution, which controls tail chain transitions from non-extreme to extreme. Although distributional convergence

---

Date: December 30, 2011.

2010 *Mathematics Subject Classification.* Primary 60G70, 60J05; Secondary 62P05.

Key words and phrases. Extreme values, multivariate regular variation, Markov chain, transition kernel, tail chain, heavy tails.

S. I. Resnick and D. Zeber were partially supported by ARO Contract W911NF-10-1-0289 and NSA Grant H98230-11-1-0193 at Cornell University.

results are more naturally phrased in terms of transition kernels, we treat the equivalent update function forms as an integral component to interfacing with applications, and we phrase relevant assumptions in terms of both. While not making explicit a complete tail chain model for the class of chains excluded previously, we demonstrate the extent to which previous models may be viewed as a partial approximation within our framework. This is accomplished by formalizing the division between extreme and non-extreme states as a level we term the *extremal boundary*. We show that, in general, the tail chain approximates the *extremal component*, the portion of the original chain having yet to cross below this boundary. Phrased in these terms, the regularity condition requires that the distinction between the original chain and its extremal component disappears asymptotically.

After introducing our extreme value theory for transition kernels, along with a representation in terms of update functions, we derive limits of finite-dimensional distributions conditional on the initial state, as it becomes extreme. We then examine the effect of the regularity condition on these results. Finally, adding the assumption of marginal regularly varying tails leads to convergence results for the unconditional distributions akin to regular variation.

1.1. **Notation and Conventions.** We review notation and relevant concepts. If not explicitly specified, assume that any space $\mathbb{S}$ under discussion is a topological space paired with its Borel σ-field of open sets $\mathcal{B}(\mathbb{S})$ to form a measurable space. Denote by $\mathcal{K}(\mathbb{S})$ the collection of its compact sets; by $\mathcal{C}(\mathbb{S})$ the space of real-valued continuous, bounded functions on $\mathbb{S}$; and by $\mathcal{C}_K^+(\mathbb{S})$ the space of non-negative continuous functions with compact support. Weak convergence of probability measures is represented by $\Rightarrow$.

For a space $\mathbb{E}$ which is locally compact with countable base (for example, a subset of $[-\infty,\infty]^d$), $\mathbb{M}_+(\mathbb{E})$ is the space of non-negative Radon measures on $\mathcal{B}(\mathbb{E})$; point measures consisting of single point masses at x will be written as $\epsilon_x(\cdot)$. A sequence of measures $\{\mu_n\} \subset \mathbb{M}_+(\mathbb{E})$ converges vaguely to $\mu \in \mathbb{M}_+(\mathbb{E})$ (written $\mu_n \xrightarrow{v} \mu$) if $\int_{\mathbb{E}} f\, d\mu_n \to \int_{\mathbb{E}} f\, d\mu$ as $n \to \infty$ for any $f \in \mathcal{C}_K^+(\mathbb{E})$. The shorthand $\mu(f) = \int f\ d\mu$ is handy. That the distribution of a random vector $\boldsymbol{X}$ is regularly varying on a cone $\mathbb{E} \subset [-\infty,\infty]^d \backslash \{\boldsymbol{0}\}$ means that $t\,\mathsf{P}[\boldsymbol{X}/b(t) \in \cdot\,] \xrightarrow{v} \mu^*(\cdot)$ in $\mathbb{M}_+(\mathbb{E})$ as $t \to \infty$ for some non-degenerate limit measure $\mu^* \in \mathbb{M}_+(\mathbb{E})$ and scaling function $b(t) \to \infty$. The limit μ^* is necessarily homogeneous in the sense that $\mu^*(c\,\cdot) = c^{-\alpha}\mu^*(\cdot)$ for some $\alpha > 0$. The regular variation is *standard* if $b(t) = t$.

If $\boldsymbol{X} = (X_0, X_1, X_2, \dots)$ is a (homogeneous) Markov chain and K is a Markov transition kernel, we write $\boldsymbol{X} \sim K$ to mean that the dependence structure of $\boldsymbol{X}$ is specified by K, i.e.

$$\mathsf{P}[X_{n+1} \in \cdot \mid X_n = x] = K\big(x\,,\,\cdot\big), \qquad n = 0, 1, \dots .$$

We adopt the standard shorthand $\mathsf{P}_x[(X_1,\dots,X_m) \in \cdot\,] = \mathsf{P}[(X_1,\dots,X_m) \in \cdot \mid X_0 = x]$. Some useful technical results are assembled in Section 8 (p. 21).

2. Extremal Theory for Markov Kernels

We begin by focusing on the Markov transition kernels rather than the stochastic processes they determine, and introduce a class of kernels we term "tail kernels," which we will view as scaling limits of certain kernels. Antecedents include Segers' [23] definition of "back-and-forth tail chains" that approximate certain Markov chains started from an extreme value.

For a Markov chain $\boldsymbol{X} \sim K$ on $[0,\infty)$, it is reasonable to expect that extremal behaviour of $\boldsymbol{X}$ is determined by pairs (X_n, X_{n+1}), and one way to control such pairs is to assume that (X_n, X_{n+1}) belongs to a bivariate domain of attraction (cf. [5, 24]). In the context of regular variation, writing

$$t\,\mathsf{P}\left[\frac{X_n}{b(t)} \in A_0\,,\, \frac{X_{n+1}}{b(t)} \in A_1\right] = \int_{A_0} K\big(b(t)u\,,\, b(t)A_1\big)\; t\,\mathsf{P}\left[\frac{X_n}{b(t)} \in du\right] \tag{2.1}$$

suggests combining marginal regular variation of X_n with a scaling kernel limit to derive extremal properties of the finite-dimensional distributions (fdds) [18, 19, 23], and this is the direction we take. We first discuss the kernel scaling operation.

For simplicity, we assume the state space of the Markov chain is $[0,\infty)$, although with suitable modifications, it is relatively straightforward to extend the results to $\mathbb{R}^d$. Henceforth G and H will denote probability distributions on $[0,\infty)$.

2.1. **Tail Kernels.** The *tail kernel associated with* G, with *boundary distribution* H, is

$$K^*\big(y\,,\,A\big) = \begin{cases} G(y^{-1}A) & y>0 \\ H(A) & y=0 \end{cases} \tag{2.2}$$

for any measurable set A. Thus, the class of tail kernels on $[0,\infty)$ is parameterized by the pair of probability distributions (G,H). Such kernels are characterized by a scaling property:

Proposition 2.1. *A Markov transition kernel K is a tail kernel associated with some (G,H) if and only if it satisfies the relation*

$$K\big(uy\,,\,A\big) = K\big(y\,,\,u^{-1}A\big) \tag{2.3}$$

when $y>0$ for any $u>0$, in which case $G(\cdot)=K(1\,,\,\cdot)$. The property (2.3) *extends to $y=0$ iff $H=\epsilon_0$.*

Proof. If K is a tail kernel, (2.3) follows directly from the definition. Conversely, assuming (2.3), for $y>0$ we can write

$$K\big(y\,,\,A\big) = K\big(1\,,\,y^{-1}A\big),$$

demonstrating that K is a tail kernel associated with $K(1\,,\,\cdot)$ (with boundary distribution $H=K(0\,,\,\cdot)$). To verify the second assertion, fixing $u>0$, we must show that $H(u^{-1}\cdot)=H(\cdot)$ iff $H=\epsilon_0$. On the one hand, we have $\epsilon_0(u^{-1}A)=\epsilon_0(A)$. On the other, $H(0,\infty)=\lim_{n\to\infty}H(n^{-1},\infty)=H(1,\infty)$, so $H(0,1]=0$. A similar argument shows that $H(1,\infty)=0$ as well. □

We call the Markov chain $\boldsymbol{T}\sim K^*$ a *tail chain associated with* (G,H). Such a chain can be represented as

$$T_n = \xi_n\,T_{n-1} + \xi_n'\,\mathbf{1}_{\{T_{n-1}=0\}} \qquad \text{for } n=1,2,\dots, \tag{2.4}$$

where $\xi_n \overset{\text{iid}}{\sim} G$ and $\xi_n' \overset{\text{iid}}{\sim} H$ are independent of each other and of T_0. If $H=\epsilon_0$, then $\boldsymbol{T}$ becomes a multiplicative random walk with step distribution G and absorbing barrier at $\{0\}$: $T_n=T_0\,\xi_1\cdots\xi_n$.

2.2. **Convergence to Tail Kernels.** The tail chain approximates the behaviour of a Markov chain $\boldsymbol{X}\sim K$ in extreme states. Asymptotic results require that the normalized distribution of X_1 be well-approximated by some distribution G when X_0 is large, and we interpret this requirement as a domain of attraction condition for kernels.

Definition. A Markov transition kernel $K:[0,\infty)\times\mathcal{B}[0,\infty)\to[0,1]$ is in the *domain of attraction of* G, written $K\in D(G)$, if as $t\to\infty$,

$$K\big(t\,,\,t\,\cdot\,\big)\Rightarrow G(\cdot) \qquad \text{on } [0,\infty]. \tag{2.5}$$

Note that $D(G)$ contains at least the class of tail kernels associated with G (i.e. with any boundary distribution H). A simple scaling argument extends (2.5) to

$$K\big(tu\,,\,t\,\cdot\,\big)\Rightarrow G(u^{-1}\cdot) =: K^*\big(u\,,\,\cdot\big), \qquad u>0, \tag{2.6}$$

where K^* is any tail kernel associated with G; this is the form appearing in (2.1). Thus tail kernels are scaling limits for kernels in a domain of attraction. In fact, tail kernels are the only possible limits:

Proposition 2.2. *Let K be a transition kernel and H be an arbitrary distribution on $[0,\infty)$. If for each $u>0$ there exists a distribution G_u such that $K\big(tu\,,\,t\,\cdot\big) \Rightarrow G_u(\cdot)$ as $t\to\infty$, then the function $\widehat{K}$ defined on $[0,\infty)\times\mathcal{B}[0,\infty)$ as*

$$\widehat{K}\big(u\,,\,A\big) := \begin{cases} G_u(A) & u>0 \\ H(A) & u=0 \end{cases}$$

is a tail kernel associated with G_1.

Proof. It suffices to show that $G_u(\cdot) = G_1(u^{-1}\cdot)$ for any $u>0$. But this follows directly from the uniqueness of weak limits, since (2.6) shows that $K(tu\,,\,t\,\cdot) \Rightarrow G_1(u^{-1}\cdot)$. □

A version of (2.6) uniform in u is needed for fdd convergence results.

Proposition 2.3. *Suppose $K\in D(G)$, and K^* is a tail kernel associated with G. Then, for any $u>0$ and any non-negative function $u_t = u(t)$ such that $u_t\to u$ as $t\to\infty$, we have*

$$K\big(tu_t\,,\,t\,\cdot\big) \Rightarrow K^*\big(u\,,\,\cdot\big), \qquad (t\to\infty). \tag{2.7}$$

Proof. Suppose $u_t\to u>0$. Observe that $K(tu_t\,,\,t\,\cdot) = K(tu_t\,,\,(tu_t)\,u_t^{-1}\cdot)$, and put $h_t(x) = u_t x$, $h(x)=ux$. Writing $P_t(\cdot) = K(tu_t\,,\,tu_t\,\cdot)$, we have

$$K\big(tu_t\,,\,t\,\cdot\big) = P_t\circ h_t^{-1} \Rightarrow G\circ h^{-1} = G(u^{-1}\cdot) = K^*\big(u\,,\,\cdot\big)$$

by [2, Theorem 5.5, p. 34]. □

The measure G controls $\boldsymbol{X}$ upon leaving an extreme state, and H describes the possibility of jumping from a non-extreme state to an extreme one. The traditional assumption (2.5) provides no information about H, and in fact (2.7) may fail if $u=0$—see Example 6.2. However, the choice of H cannot be ignored if 0 is an accessible point of the state space, especially for cases where $G(\{0\}) = K^*(y\,,\,\{0\}) > 0$. We propose pursuing implications of the traditional assumption (2.5) alone, and will add conditions as needed to understand boundary behaviour of $\boldsymbol{X}$.

Alternative, more general formulations of (2.5) include replacing $K(t\,,\,t\,\cdot)$ with $K(t\,,\,a(t)\,\cdot)$ or $K(t\,,\,a(t)\,\cdot + b(t))$ with appropriate functions $a(t)>0$ and $b(t)$, in analogy with the usual domains of attraction conditions in extreme value theory. Indeed, the second choice coincides with the original presentation by Perfekt [18], and relates to the conditional extreme value model [8, 13, 14]. For clarity, and to maintain ties with regular variation, we retain the standard choice $a(t)=t$, $b(t)=0$.

2.3. **Representation.** How do we characterize kernels belonging to $D(G)$? From (2.4), for chains transitioning according to a tail kernel, the next state is a random multiple of the previous one, provided the prior state is non-zero. We expect that chains transitioning according to $K\in D(G)$ behave approximately like this upon leaving a large state, and this is best expressed in terms of a function describing how a new state depends on the prior one.

Given a kernel K, we can always find a sample space $\mathbb{E}$, a measurable function $\psi:[0,\infty)\times\mathbb{E}\to[0,\infty)$ and an $\mathbb{E}$-valued random element V such that $\psi(y,V)\sim K(y\,,\,\cdot)$ for all y. Given a random variable X_0, if we define the process $\boldsymbol{X} = (X_0, X_1, X_2, \dots)$ recursively as

$$X_{n+1} = \psi(X_n, V_{n+1}), \qquad n\geq 0,$$

where $\{V_n\}$ is an iid sequence equal in distribution to V and independent of X_0, then $\boldsymbol{X}$ is a Markov chain with transition kernel K. Call the function ψ an *update function corresponding to K*. If in addition $K\in D(G)$, the domain of attraction condition (2.5) becomes

$$t^{-1}\psi(t,V) \Rightarrow \xi,$$

where $\xi \sim G$. Applying the probability integral transform or the Skorohod representation theorems [3, Theorem 3.2, p. 6], [4, Theorem 6.7, p. 70], we get the following result.

Proposition 2.4. *If K is a transition kernel, $K \in D(G)$ if and only if there exists a measurable function $\psi^* : [0,\infty) \times [0,1] \to [0,\infty)$ and a random variable $\xi^* \sim G$ on the uniform probability space $([0,1], \mathcal{B}, \lambda)$ such that*

$$t^{-1}\psi^*(t,u) \longrightarrow \xi^*(u) \qquad \forall\, u \in [0,1] \tag{2.8}$$

as $t \to \infty$, and ψ^* is an update function corresponding to K in the sense that

$$\lambda\big[\psi^*(y,\cdot) \in A\,\big] = K\big(y\,,\,A\big)$$

for measurable sets A.

Think of the update function as $\psi^*(y,U)$ where $U(u) = u$ is a uniform random variable on $[0,1]$.

Proof. If there exist such ψ^* and ξ^* satisfying (2.8) then clearly $K \in D(G)$. Conversely, suppose $\psi(\cdot, V)$ is an update function corresponding to K. According to Skorohod's representation theorem (cf. Billingsley [4] p. 70, with the necessary modifications to allow for an uncountable index set), there exists a random variable ξ^* and a stochastic process $\{Y_t^*\,;\, t \geq 0\}$ defined on the uniform probability space $([0,1], \mathcal{B}, \lambda)$, taking values in $[0,\infty)$, such that

$$\xi^* \sim G\,, \qquad Y_0^* \overset{d}{=} \psi(0,V)\,, \qquad Y_t^* \overset{d}{=} t^{-1}\psi(t,V) \ \text{ for } t > 0,$$

and $Y_t^*(u) \to \xi^*(u)$ as $t \to \infty$ for every $u \in [0,1]$. Now, define $\psi^* : [0,\infty) \times [0,1] \to [0,\infty)$ as

$$\psi^*(0,u) = Y_0^*(u) \qquad \text{and} \qquad \psi^*(t,u) = tY_t^*(u)\,, \quad t > 0, \qquad \forall\, u \in [0,1].$$

It is evident that $\lambda[\psi^*(y,\cdot) \in A] = \mathsf{P}[\psi(y,V) \in A]$ for $y \in [0,\infty)$, so ψ^* is indeed an update function corresponding to K, and ψ^* satisfies (2.8) by construction. □

Update functions corresponding to K are not unique, and some of them may fail to converge pointwise as in (2.8). However (2.8) is convenient, and Proposition 2.4 shows that Segers' [23] Condition 2.2 in terms of update functions is equivalent to our weak convergence formulation $K \in D(G)$.

Pointwise convergence in (2.8) gives an intuitive representation of kernels in a domain of attraction.

Corollary 2.1. *$K \in D(G)$ iff there exists a random variable $\xi \sim G$ defined on the uniform probability space, and a measurable function $\phi : [0,\infty) \times [0,1] \to (-\infty,\infty)$ satisfying $t^{-1}\phi(t,u) \to 0$ for all $u \in [0,1]$ such that*

$$\psi(y,u) := \xi(u)\, y + \phi(y,u) \tag{2.9}$$

is an update function corresponding to K.

Proof. If such ξ and ϕ exist, then $t^{-1}\psi(t,u) = \xi(u) + t^{-1}\phi(t,u) \to \xi(u)$ for all u, so ψ satisfies (2.8). The converse follows from (2.8). □

Many Markov chains such as ARCH, GARCH and autoregressive processes are specified by structured recursions that allow quick recognition of update functions corresponding to kernels in a domain of attraction. A common example is the update function $\psi(y,(Z,W)) = Zy + W$, which behaves like $\psi'(y,Z) = Zy$ when y is large—compare ψ' to the form (2.4) discussed for tail kernels. In general, if K has an update function ψ of the form

$$\psi(y,(Z,W)) = Zy + \phi(y,W) \tag{2.10}$$

for a random variable $Z \geq 0$ and a random element W, where $t^{-1}\phi(t,w) \to 0$ whenever $w \in C$ for which $\mathsf{P}[W \in C] = 1$, then $K \in D(G)$ with $G = \mathsf{P}[Z \in \cdot\,]$. We will refer to update functions satisfying (2.10) as being in *canonical form*.

3. Finite-Dimensional Convergence and the Extremal Component

Given a Markov chain $\boldsymbol{X} \sim K \in D(G)$, we show that the finite-dimensional distributions (fdds) of $\boldsymbol{X}$, started from an extreme state, converge to those of the tail chain $\boldsymbol{T}$ defined in (2.4). We initially develop results that depend only on G (but not H), and then clarify what behaviour of $\boldsymbol{X}$ is controlled by G and H respectively. We make explicit links with prior work that did not consider the notion of boundary distribution.

If $G(\{0\}) = 0$, the choice of H is inconsequential, since $\mathsf{P}\big[\boldsymbol{T}$ eventually hits $\{0\}\big] = 0$ and $\boldsymbol{T}$ is indistinguishable from the multiplicative random walk $\{T_n^* = T_0\xi_1 \cdots \xi_n, n \geq 0\}$ (where $T_0 > 0$ and $\{\xi_n\}$ are iid $\sim G$ and independent of T_0). In this case, assume without loss of generality that $H = \epsilon_0$. However, if $G(\{0\}) > 0$, any result not depending on H must be restricted to fdds conditional on the tail chain not having yet hit $\{0\}$. For example, consider the trajectory of $(X_1, \dots, X_m)$, started from $X_0 = t$, through the region $(t, \infty)^{m-2} \times [0, \delta] \times (t, \infty)$, where t is a high level. The tail chain would model this as a path through $(0, \infty)^{m-2} \times \{0\} \times (0, \infty)$, which requires specifying H to control transitions away from $\{0\}$.

This raises the question of how to interpret the first hitting time of $\{0\}$ for $\boldsymbol{T}$ in terms of the original Markov chain $\boldsymbol{X}$. Such hitting times are important in the study of Markov chain point process models of exceedance clusters based on the tail chain. Intuitively, a transition to $\{0\}$ by $\boldsymbol{T}$ represents a transition from an extreme state to a non-extreme state by $\boldsymbol{X}$. We make this notion precise in Section 3.2 by viewing such transitions as downcrossings of a certain level we term the "extremal boundary."

We assume $\boldsymbol{X}$ is a Markov chain on $[0, \infty)$ with transition kernel $K \in D(G)$, K^* is a tail kernel associated with G with unspecified boundary distribution H, and $\boldsymbol{T}$ is a Markov chain on $[0, \infty)$ with kernel K^*. The finite-dimensional distributions of $\boldsymbol{X}$, conditional on $X_0 = y$, are given by

$$\mathsf{P}_y\big[(X_1, \dots, X_m) \in d\boldsymbol{x}_m\big] = K\big(y\,,\, dx_1\big) K\big(x\,,\, dx_2\big) \cdots K\big(x_{m-1}\,,\, dx_m\big),$$

and analogously for $\boldsymbol{T}$.

3.1. FDDs Conditional on the Intial State.

Define the conditional distributions

$$\pi_m^{(t)}\big(u\,,\,\cdot\big) = \mathsf{P}_{tu}\left[\left(\frac{X_1}{t}, \dots, \frac{X_m}{t}\right) \in \cdot\,\right] \quad \text{and} \quad \pi_m\big(u\,,\,\cdot\big) = \mathsf{P}_u\big[(T_1, \dots, T_m) \in \cdot\,\big], \quad m \geq 1, \tag{3.1}$$

on $[0, \infty) \times \mathcal{B}[0, \infty]^m$. We consider when $\pi_m^{(t)} \Rightarrow \pi_m$ on $[0, \infty]^m$ pointwise in u. If $G(\{0\}) = 0$, this is a direct consequence of the domain of attraction condition (2.5), but if $G(\{0\}) > 0$, more thought is required. We begin by restricting the convergence to the smaller space $\mathbb{E}'_m := (0, \infty]^{m-1} \times [0, \infty]$. Relatively compact sets in $\mathbb{E}'_m$ are contained in rectangles $[\boldsymbol{a}, \boldsymbol{\infty}] \times [0, \infty]$, where $\boldsymbol{a} \in (0, \infty)^{m-1}$.

Theorem 3.1. *Let $u_t = u(t)$ be a non-negative function such that $u_t \to u > 0$ as $t \to \infty$.*

(a) *The restrictions to $\mathbb{E}'_m$,*

$$\mu_m^{(t)}\big(u\,,\,\cdot\big) := \pi_m^{(t)}\big(u\,,\,\cdot \cap \mathbb{E}'_m\big) \quad \textit{and} \quad \mu_m\big(u\,,\,\cdot\big) := \pi_m\big(u\,,\,\cdot \cap \mathbb{E}'_m\big), \tag{3.2}$$

satisfy

$$\mu_m^{(t)}\big(u_t\,,\,\cdot\big) \xrightarrow{v} \mu_m\big(u\,,\,\cdot\big) \qquad \textit{in } \mathbb{M}_+(\mathbb{E}'_m) \qquad (t \to \infty). \tag{3.3}$$

(b) *If $G(\{0\}) = 0$, we have*

$$\pi_m^{(t)}\big(u_t\,,\,\cdot\big) \Rightarrow \pi_m\big(u\,,\,\cdot\big) \qquad \textit{on } [0, \infty]^m \qquad (t \to \infty). \tag{3.4}$$

Proof. The Markov structure suggests an induction argument facilitated by Lemma 8.2 (p. 21). Consider (a) first. If $m = 1$, then (3.3) above reduces to (2.7). Assume $m \geq 2$, and let $f \in \mathcal{C}_K^+(\mathbb{E}'_m)$.

Writing $\mathbb{E}'_m = (0,\infty] \times \mathbb{E}'_{m-1}$, we can find $a > 0$ and $B \in \mathcal{K}(\mathbb{E}'_{m-1})$ such that f is supported on $[a,\infty] \times B$. Now, observe that

$$\begin{aligned}\mu_m^{(t)}\big(u_t\,,\,\cdot\big)(f) &= \int_{(0,\infty]} K\big(tu_t\,,\,tdx_1\big) \int_{\mathbb{E}'_{m-1}} K\big(tx_1\,,\,tdx_2\big)\cdots K\big(tx_{m-1}\,,\,tdx_m\big)\; f(\boldsymbol{x}_m) \\ &= \int_{(0,\infty]} K\big(tu_t\,,\,tdx_1\big) \int_{\mathbb{E}'_{m-1}} \mu_{m-1}^{(t)}\big(x_1\,,\,d(x_2,\ldots,x_m)\big)\; f(\boldsymbol{x}_m).\end{aligned}$$

Defining

$$h_t(v) = \int_{\mathbb{E}'_{m-1}} \mu_{m-1}^{(t)}\big(v\,,\,d\boldsymbol{x}_{m-1}\big)\; f(v,\boldsymbol{x}_{m-1}) \quad \text{and} \quad h(v) = \int_{\mathbb{E}'_{m-1}} \mu_{m-1}\big(v\,,\,d\boldsymbol{x}_{m-1}\big)\; f(v,\boldsymbol{x}_{m-1})\,,$$

the previous expression becomes

$$\mu_m^{(t)}\big(u_t\,,\,\cdot\big)(f) = \int_{(0,\infty]} K\big(tu_t\,,\,tdv\big)\; h_t(v).$$

Now, suppose $v_t \to v > 0$: we verify

$$h_t(v_t) \longrightarrow h(v). \tag{3.5}$$

By continuity, we have $f(v_t,\boldsymbol{x}^t_{m-1}) \to f(v,\boldsymbol{x}_{m-1})$ whenever $\boldsymbol{x}^t_{m-1} \to \boldsymbol{x}_{m-1}$, and the induction hypothesis provides $\mu_{m-1}^{(t)}(v_t\,,\,\cdot) \xrightarrow{v} \mu_{m-1}(v\,,\,\cdot)$. Also, $f(x,\cdot)$ has compact support B (without loss of generality, $\mu_{m-1}(v\,,\,\partial B) = 0$). Combining these facts, (3.5) follows from Lemma 8.2 (b). Next, since the h_t and h have common compact support $[a,\infty]$, and recalling from Propostion 2.3 that $K(tu_t\,,\,t\,\cdot) \Rightarrow K^*(u\,,\,\cdot)$, Lemma 8.2 (a) yields

$$\mu_m^{(t)}\big(u_t\,,\,\cdot\big)(f) \longrightarrow \int_{(0,\infty]} K^*\big(u\,,\,dv\big)\; h(v) = \mu_m\big(u\,,\,\cdot\big)(f).$$

Implication (b) follows from essentially the same argument. For $m \geq 2$, suppose $f \in \mathcal{C}[0,\infty]^m$. Replacing μ by π and $\mathbb{E}'_{m-1}$ by $[0,\infty]^{m-1}$ in the definitions of h_t and h, we have

$$\pi_m^{(t)}\big(u_t\,,\,\cdot\big)(f) = \int_{[0,\infty]} K\big(tu_t\,,\,tdv\big)\; h_t(v).$$

This time Lemma 8.2 (a) shows that $h_t(v_t) \to h(v)$ if $v_t \to v > 0$, and since $K^*(u\,,\,(0,\infty]) = 1$, resorting to Lemma 8.2 (a) once more yields

$$\pi_m^{(t)}\big(u_t\,,\,\cdot\big)(f) \longrightarrow \int_{[0,\infty]} K^*\big(u\,,\,dv\big)\; h(v) = \pi_m\big(u\,,\,\cdot\big)(f). \qquad \square$$

If $G(\{0\}) > 0$, then $K^*(u\,,\,(0,\infty]) = 1 - G(\{0\}) < 1$, and for (3.4) to hold would require knowing the behaviour of $h_t(v_t)$ when $v_t \to 0$ as well. Behaviour near zero is controlled by an asymptotic condition related to the boundary distribution H. Previous work handled this using the regularity condition discussed in Section 4.

3.2. **The Extremal Boundary.** The normalization employed in the domain of attraction condition (2.5) suggests that, starting from a large state t, the extreme states are approximately scalar multiples of t. For example, we would consider a transition from t into $(t/3, 2t]$ to remain extreme. Thus, we think of states which can be made smaller than $t\delta$ for any δ, if t is large enough, as non-extreme. In this context, the set $[0,\sqrt{t}]$ would consist of non-extreme states.

Under (2.5), a tail chain path through $(0,\infty)$ models the original chain $\boldsymbol{X}$ travelling among extreme states, and all of the non-extreme states are compacted into the state $\{0\}$ in the state space of $\boldsymbol{T}$. Therefore, if $\boldsymbol{X}$ is started from an extreme state, the portion of the tail chain depending solely on G is informative up until the first time $\boldsymbol{X}$ crosses down to a non-extreme state. If $G(\{0\}) = 0$,

such a transition would become more and more unlikely as the initial state increases in which case G provides a complete description of the behaviour of $\boldsymbol{X}$ in any finite number of steps following a visit to an extreme state (Theorem 3.1 (b)).

Drawing upon this interpretation, we develop a rigorous formulation of the distinction between extreme and non-extreme states, and we recast Theorem 3.1 as convergence on the unrestricted space $[0,\infty]^m$ of the conditional fdds, given that $\boldsymbol{X}$ has not yet reached a non-extreme state.

Definition. Suppose $K \in D(G)$. An *extremal boundary* for K is a non-negative function $y(t)$ defined on $[0,\infty)$, satisfying $\lim t \to \infty y(t) = 0$ and

$$K\big(t\,,\, t\,[0,y(t)]\big) \longrightarrow G(\{0\}) \qquad \text{as} \;\; t \to \infty. \tag{3.6}$$

Such a function is guaranteed to exist by Lemma 8.5 (p. 23).

If $G(\{0\}) = 0$, then $y(t) \equiv 0$ is a trivial choice. For any function $0 \le y(t) \to 0$, we have $\limsup_{t\to\infty} K(t\,,\, t\,[0,y(t)]) \le G(\{0\})$, so (3.6) is equivalent to

$$\liminf_{t\to\infty} K\big(t\,,\, t\,[0,y(t)]\big) \ge G(\{0\}). \tag{3.7}$$

If $y(t)$ is an extremal boundary, it follows that any function $0 \le \tilde{y}(t) \to 0$ with $\tilde{y}(t) \ge y(t)$ for $t \ge t_0$ is also an extremal boundary for K. Taking $\tilde{y}(t) = \vee_{s\ge t}\, y(s)$ shows that without loss of generality, we can assume $y(t)$ to be non-increasing.

The extremal boundary has a natural formulation in terms of the update function. As in (2.10), let $\psi(y,(Z,W)) = Zy + \phi(y,W)$ be an update function in canonical form, where y is extreme. If $Z > 0$ then the next state is approximately Zy, another extreme state. Otherwise, if $Z = 0$, the next state is $\phi(y,W)$, and a transition from an extreme to a non-extreme state has taken place. This suggests choosing an extremal boundary whose order is between t and $\phi(t,w)$.

Proposition 3.1. *Suppose $\psi(y,(Z,W))$ is an update function in canonical form as in* (2.10)*. If $\zeta(t) > 0$ is a function on $[0,\infty)$ such that*

$$\phi(t,w)/\zeta(t) \longrightarrow 0 \tag{3.8}$$

as $t \to \infty$ whenever $w \in B$ for which $\mathsf{P}[W \in B] = 1$, then $\liminf_{t\to\infty} K(t\,,\,[0,\zeta(t)]) \ge G(\{0\})$. Provided $\lim_{t\to\infty} \zeta(t)/t = 0$, an extremal boundary is given by $y(t) := \zeta(t)/t$.

Thus if $\phi(t,w) = o(\zeta(t))$ and $\zeta(t) = o(t)$ then $\zeta(t)/t$ is an extremal boundary. For example, if $\psi(y,(Z,W)) = Zy + W$, so that $\phi(t,w) = w$, then choosing $\zeta(t)$ to be any function $\zeta(t) \to \infty$ such that $\zeta(t) = o(t)$ makes $\zeta(t)/t$ an extremal boundary. Choosing $\zeta(t) = \sqrt{t}$, we find that $y(t) = 1/\sqrt{t}$ is an extremal boundary.

Proof. Since

$$\begin{aligned}\mathsf{P}\big[\psi(t) \le \zeta(t)\,,\, Z=0\big] &= \mathsf{P}\big[\phi(t,W) \le \zeta(t)\,,\, Z=0\big] \ge \mathsf{P}\big[|\phi(t,W)| \le \zeta(t)\,,\, Z=0\big] \\ &\ge \mathsf{P}\big[Z=0\big] - \mathsf{P}\left[\frac{|\phi(t,W)|}{\zeta(t)} > 1\right] \longrightarrow \mathsf{P}\big[Z=0\big],\end{aligned}$$

we have

$$\liminf_{t\to\infty} K\big(t\,,\,[0,\zeta(t)]\big) = \liminf_{t\to\infty} \mathsf{P}\big[\psi(t) \le \zeta(t)\big] \ge \mathsf{P}\big[Z=0\big]. \qquad \square$$

We will need an extremal boundary for which (3.6) still holds upon replacing the initial state t with tu_t, where $u_t \to u > 0$. Compare the following extension with Proposition 2.3.

Proposition 3.2. *If $K \in D(G)$, then there exists an extremal boundary $y^*(t)$ such that*

$$K\big(tu_t\,,\, t\,[0,y^*(t)]\big) \longrightarrow G(\{0\}) \qquad \textit{as} \;\; t \to \infty \tag{3.9}$$

for any non-negative function $u_t = u(t) \to u > 0$.

We will refer to y^* as a *uniform extremal boundary.*

Proof. Let $y(t)$ be an extremal boundary for K. As a first step, fix $u_0 > 1$, and suppose $u_0^{-1} < u < u_0$. Define $\tilde{y}(t) = u_0\, y(tu_0^{-1})$. Now, if $u_t \to u$, then $y_{\{u\}}(t) := u_t\, y(tu_t)$ satisfies (3.9), since

$$K\big(tu_t\,,\, t\,[0, y_{\{u\}}(t)]\big) = K\big(tu_t\,,\, tu_t\,[0, y(tu_t)]\big) \longrightarrow G(\{0\}).$$

Here $y_{\{u\}}$ depends on the choice of function u_t. However, since we eventually have $u_0^{-1} < u_t < u_0$ for t large enough, it follows that $\tilde{y}(t) > y_{\{u\}}(t)$ for such t. Hence, $\tilde{y}(t)$ satisfies (3.9) for any $u_t \to u$ with $u_0^{-1} < u < u_0$.

Next, we remove the restriction in u_0 via a diagonalization argument. For $k = 2, 3, \ldots$, let $y_k(t)$ be extremal boundaries such that $K(tu_t\,,\, t\,[0, y_k(t)]) \to G(\{0\})$ whenever $u_t \to u$ for $u \in (k^{-1}, k)$, and put $y_0 = y_1 = y$. Next, define the sequence $\{(s_k, x_k) : k = 0, 1, \ldots\}$ inductively as follows. Setting $s_0 = 0$ and $x_0 = y_0(1)$, choose $s_k \geq s_{k-1} + 1$ such that $y_j(t) \leq k^{-1} \wedge x_{k-1}$ for all $j = 0, \ldots, k$ whenever $t \geq s_k$, and put $x_k = \max\{y_j(s_k) : j = 0, \ldots, k\}$. Note that $x_k \leq k^{-1} \wedge x_{k-1}$, so $x_k \downarrow 0$, and $s_k \uparrow \infty$. Finally, set

$$y^*(t) = \sum_{k=0}^{\infty} x_k\, \mathbf{1}_{[s_k,\, s_{k+1})}(t).$$

Observe that $0 \leq y^*(t) \downarrow 0$, and suppose $u_t \to u > 0$. Then $u \in (k_0^{-1}, k_0)$ for some k_0, so $K(tu_t\,,\, t\,[0, y_{k_0}(t)]) \to G(\{0\})$, and for $k \geq k_0$, our construction ensures that whenever $s_k \leq t < s_{k+1}$, we have $y_{k_0}(t) \leq y_{k_0}(s_k) \leq x_k = y^*(t)$. Therefore, $y^*(t) \geq y_{k_0}(t)$ for $t \geq s_{k_0}$, so y^* satisfies (3.9). ◻

Henceforth, we assume any $K \in D(G)$ is accompanied by a uniform extremal boundary denoted by $y(t)$, and we consider extreme states on the order of t to be $(ty(t), \infty]$. If $G(\{0\}) = 0$, then all positive states are extreme states. We now use the extremal boundary to reformulate the convergence of Theorem 3.1 on the larger space $[0, \infty]^m$. Put $\mathbb{E}'_m(t) = (y(t), \infty]^{m-1} \times [0, \infty]$, so that $\mathbb{E}'_m(t) \uparrow \mathbb{E}'_m = (0, \infty]^{m-1} \times [0, \infty]$. Recall the notation $\mu_m^{(t)}$ and μ_m^* from (3.1), (3.2) in Theorem 3.1 (p. 6).

Theorem 3.2. *Let $u_t = u(t)$ be a non-negative function such that $u_t \to u > 0$ as $t \to \infty$. Taking*

$$\widetilde{\mu}_m^{(t)}\big(u\,,\,\cdot\,\big) = \pi_m^{(t)}\big(u\,,\,\cdot \cap \mathbb{E}'_m(t)\big),$$

we have

$$\widetilde{\mu}_m^{(t)}\big(u_t\,,\,\cdot\,\big) \xrightarrow{v} \mu_m\big(u\,,\,\cdot\,\big) \qquad \text{in } \mathbb{M}_+[0,\infty]^m \qquad (t \to \infty).$$

Proof. Note that we can just as well write $\widetilde{\mu}_m^{(t)}(u\,,\,\cdot) = \mu_m^{(t)}(u\,,\,\cdot \cap \mathbb{E}'_m(t))$. Suppose $m \geq 2$ and let $f \in \mathcal{C}_K^+[0,\infty]^m$. For $\delta > 0$, define $A_\delta = (\delta, \infty]^{m-1} \times [0, \infty]$, and choose δ such that $\mu_m(u\,,\,\partial A_\delta) = 0$. On the one hand, for large t we have

$$\begin{aligned} \widetilde{\mu}_m^{(t)}\big(u_t\,,\,\cdot\big)(f) = \int_{[0,\infty]^m} f(\boldsymbol{x})\, \mathbf{1}_{\mathbb{E}'_m(t)}(\boldsymbol{x})\; \mu_m^{(t)}\big(u_t\,,\, d\boldsymbol{x}\big) &\geq \int_{\mathbb{E}'_m} f(\boldsymbol{x})\, \mathbf{1}_{A_\delta}(\boldsymbol{x})\; \mu_m^{(t)}\big(u_t\,,\, d\boldsymbol{x}\big) \\ \longrightarrow \int_{\mathbb{E}'_m} f(\boldsymbol{x})\, \mathbf{1}_{A_\delta}(\boldsymbol{x})\; \mu_m\big(u\,,\, d\boldsymbol{x}\big) \end{aligned}$$

as $t \to \infty$ by Lemma 8.3 (p. 22). Letting $\delta \downarrow 0$ yields

$$\liminf_{t\to\infty}\; \widetilde{\mu}_m^{(t)}\big(u_t\,,\,\cdot\big)(f) \;\geq\; \mu_m\big(u\,,\,\cdot\big)(f) \tag{3.10}$$

by monotone convergence. On the other hand, fixing δ, we can decompose the space according to the first downcrossing of δ:

$$\widetilde{\mu}_m^{(t)}\big(u_t\,,\,\cdot\big)(f) = \int_{[0,\infty]^m} f(\boldsymbol{x})\,\mathbf{1}_{A_\delta}(\boldsymbol{x})\;\widetilde{\mu}_m^{(t)}\big(u_t\,,\,d\boldsymbol{x}\big) + \sum_{k=1}^{m-1}\int_{[0,\infty]^m} f(\boldsymbol{x})\,\mathbf{1}_{A_\delta^k}(\boldsymbol{x})\;\widetilde{\mu}_m^{(t)}\big(u_t\,,\,d\boldsymbol{x}\big), \tag{3.11}$$

where $A_\delta^k = (\delta,\infty]^{k-1}\times[0,\delta]\times[0,\infty]^{m-k}$. On the subsets A_δ^k we appeal to the bound on f, say M, to obtain

$$\int_{[0,\infty]^m} f(\boldsymbol{x})\,\mathbf{1}_{A_\delta^k}(\boldsymbol{x})\;\widetilde{\mu}_m^{(t)}\big(u_t\,,\,d\boldsymbol{x}\big) \le M\,\widetilde{\mu}_m^{(t)}\big(u_t\,,\,A_\delta^k\big).$$

Now,

$$\begin{aligned}\widetilde{\mu}_m^{(t)}\big(u_t\,,\,A_\delta^k\big) &\le \mu_k^{(t)}\big(u_t\,,\,(\delta,\infty]^{k-1}\times(y(t),\delta]\big)\\ &= \mu_k^{(t)}\big(u_t\,,\,(\delta,\infty]^{k-1}\times[0,\delta]\big) - \mu_k^{(t)}\big(u_t\,,\,(\delta,\infty]^{k-1}\times[0,y(t)]\big).\end{aligned} \tag{3.12}$$

Considering the second term, we have

$$\begin{aligned}&\mu_k^{(t)}\big(u_t\,,\,(\delta,\infty]^{k-1}\times[0,y(t)]\big)\\ &= \int_{[0,\infty]} K\big(tu_t\,,\,tdx_1\big)\mathbf{1}_{(\delta,\infty]}(x_1)\cdots\int_{[0,\infty]} K\big(tx_{k-2}\,,\,tdx_{k-1}\big)\mathbf{1}_{(\delta,\infty]}(x_{k-1})\;K\big(tx_{k-1}\,,\,t\,[0,y(t)]\big)\\ &= \int_{\mathbb{E}'_{k-1}} \mu_{k-1}^{(t)}\big(u_t\,,\,d\boldsymbol{x}_{k-1}\big)\;h_t(\boldsymbol{x}_{k-1}),\end{aligned}$$

where

$$h_t(\boldsymbol{x}_{k-1}) = K\big(tx_{k-1}\,,\,t\,[0,y(t)]\big)\,\mathbf{1}_{(\delta,\infty]^{k-1}}(\boldsymbol{x}_{k-1}).$$

Moreover, if $\boldsymbol{x}_{k-1}^t\to\boldsymbol{x}_{k-1}\in(\delta,\infty]^{k-1}$, then

$$h_t(\boldsymbol{x}_{k-1}^t) = K\big(tx_{k-1}^t\,,\,t\,[0,y(t)]\big)\,\mathbf{1}_{(\delta,\infty]^{k-1}}(\boldsymbol{x}_{k-1}^t) \longrightarrow G(\{0\})\,\mathbf{1}_{(\delta,\infty]^{k-1}}(\boldsymbol{x}_{k-1}),$$

using the fact that $y(t)$ is a uniform extremal boundary. Since $\mu_{k-1}(u\,,\,\partial(\delta,\infty]^{k-1}) = 0$ without loss of generality by choice of δ, we conclude that

$$\mu_k^{(t)}\big(u_t\,,\,(\delta,\infty]^{k-1}\times[0,y(t)]\big) \longrightarrow G(\{0\})\cdot\mu_{k-1}\big(u\,,\,(\delta,\infty]^{k-1}\big) = \mu_k\big(u\,,\,(\delta,\infty]^{k-1}\times\{0\}\big)$$

as $t\to\infty$. Now, let us return to (3.12). Given any $\epsilon>0$, by choosing δ small enough, we can make

$$\begin{aligned}\mu_k^{(t)}\big(u_t\,,\,(\delta,\infty]^{k-1}\times(y(t),\delta]\big) &\longrightarrow \mu_k\big(u\,,\,(\delta,\infty]^{k-1}\times[0,\delta]\big) - \mu_k\big(u\,,\,(\delta,\infty]^{k-1}\times\{0\}\big)\\ &\le \mu_k\big(u\,,\,(0,\infty]^{k-1}\times[0,\delta]\big) - \mu_k\big(u\,,\,(\delta,\infty]^{k-1}\times\{0\}\big)\\ &< \mu_k\big(u\,,\,(0,\infty]^{k-1}\times\{0\}\big) + \frac{\epsilon}{2} - \left(\mu_k\big(u\,,\,(0,\infty]^{k-1}\times\{0\}\big) - \frac{\epsilon}{2}\right) = \epsilon,\end{aligned}$$

i.e.

$$\limsup_{t\to\infty}\;\widetilde{\mu}_m^{(t)}\big(u_t\,,\,A_\delta^k\big) < \epsilon, \tag{3.13}$$

for $k=1,\dots,m-1$. Therefore, (3.11) implies that, given $\epsilon'>0$,

$$\begin{aligned}\limsup_{t\to\infty}\;\widetilde{\mu}_m^{(t)}\big(u_t\,,\,\cdot\,\big)(f) &\le \int_{[0,\infty]^m} f(\boldsymbol{x})\,\mathbf{1}_{A_\delta}(\boldsymbol{x})\;\mu_m\big(u\,,\,d\boldsymbol{x}\big) + M\sum_{k=1}^{m-1}\limsup_{t\to\infty}\;\widetilde{\mu}_m^{(t)}\big(u_t\,,\,A_\delta^k\big)\\ &< \mu_m\big(u\,,\,\cdot\big)(f) + \epsilon'\end{aligned}$$

for small enough δ. Combining this with (3.10) yields the result. $\square$

3.3. **The Extremal Component.** Having thus formalized the distinction between extreme and non-extreme states, we return to the question of phrasing a fdd limit result for $\boldsymbol{X}$ when H is unspecified. The extremal boundary allows us to interpret the first hitting time of $\{0\}$ by the tail chain as approximating the time of the first transition from extreme down to non-extreme. In this terminology, Theorem 3.2 provides a result, given that such a transition has yet to occur.

Define the first hitting time of a non-extreme state

$$\tau(t) = \inf\left\{n \geq 0 \,:\, X_n \leq ty(t)\right\}.$$

For a Markov chain started from tu_t, where $u_t \to u > 0$, we have $tu_t > y(t)$ for large t, so $\tau(t)$ is the first downcrossing of the extremal boundary.

For the tail chain $\boldsymbol{T}$, put $\tau^* = \inf\{n \geq 0 : T_n = 0\}$. Given $T_0 > 0$, write $\tau^* = \inf\{n \geq 1 : \xi_n = 0\}$, where $\{\xi_n\} \sim G$ are iid and independent of T_0, i.e. τ^* follows a Geometric distribution with parameter $p = G(\{0\})$. Thus, $\mathsf{P}[\tau^* = m] = p(1-p)^{m-1}$ for $m \geq 1$ if $p > 0$, and $\mathsf{P}[\tau^* = \infty] = 1$ if $p = 0$. Theorem 3.2 becomes

$$\mathsf{P}_{tu_t}\big[t^{-1}\boldsymbol{X}_m \in \cdot\,,\ \tau(t) \geq m\big] \xrightarrow{v} \mathsf{P}_u\big[\boldsymbol{T}_m \in \cdot\,,\ \tau^* \geq m\big], \tag{3.14}$$

implying that τ^* approximates $\tau(t)$:

$$\mathsf{P}_{tu_t}\big[\tau(t) \in \cdot\,\big] \Rightarrow \mathsf{P}\big[\tau^* \in \cdot\,\big], \quad (t \to \infty,\ u_t \to u > 0). \tag{3.15}$$

So if $G(\{0\}) > 0$, $\boldsymbol{X}$ takes an average of approximately $G(\{0\})^{-1}$ steps to return to a non-extreme state. but if $G(\{0\}) = 0$, $\mathsf{P}_{tu_t}[\tau_1 \leq m] \to 0$ for any $m \geq 1$ so starting from a larger and larger initial state, it will take longer and longer for $\boldsymbol{X}$ to cross down to a non-extreme state.

Let $\boldsymbol{T}^*$ be the tail chain associated with (G, ϵ_0). For $\{\xi_n\} \sim G$ iid and independent of T_0^*,

$$T_n^* = T_0^*\, \xi_1 \cdots \xi_n. \tag{3.16}$$

We restate (3.14) in terms of a process derived from $\boldsymbol{X}$, called the *extremal component* of $\boldsymbol{X}$, whose fdds converge weakly to those of $\boldsymbol{T}^*$. The extremal component is the part of $\boldsymbol{X}$ whose asymptotic behavior is controlled by G alone.

Definition. The *extremal component* of $\boldsymbol{X}$ relative to t is the process $\boldsymbol{X}^{(t)}$ defined for $t > 0$ as

$$X_n^{(t)} = X_n \cdot \mathbf{1}_{\{n < \tau(t)\}}\,, \qquad n = 0, 1, \ldots.$$

Observe that $\boldsymbol{X}^{(t)}$ is a Markov chain on $[0, \infty)$ with transition kernel

$$K^{(t)}\big(x\,,\ A\big) = \begin{cases} K\big(x\,,\ A \cap (ty(t), \infty]\big) + \epsilon_0(A) \cdot K\big(x\,,\ [0, ty(t)]\big) & x > ty(t) \\ \epsilon_0(A) & x \leq ty(t) \end{cases}.$$

It follows that $K^{(t)}(t\,,\, t\,\cdot) \Rightarrow G$ as $t \to \infty$, and additionally that $K^{(t)}(t\,,\,\{0\}) \to G(\{0\})$.

The relation between the component processes $\boldsymbol{X}^{(t)}$, $\boldsymbol{T}^*$ and the complete ones is

$$\mathsf{P}_{tu_t}\big[t^{-1}\boldsymbol{X}_m^{(t)} \in \cdot \,\big|\, \tau(t) > m\big] = \mathsf{P}_{tu_t}\big[t^{-1}\boldsymbol{X}_m \in \cdot \,\big|\, \tau(t) > m\big]$$

and

$$\mathsf{P}_u\big[\boldsymbol{T}_m^* \in \cdot \,\big|\, \tau^* > m\big] = \mathsf{P}_u\big[\boldsymbol{T}_m \in \cdot \,\big|\, \tau^* > m\big].$$

Theorem 3.3. *Let $u_t = u(t) \geq 0$ satisfy $u_t \to u > 0$ as $t \to \infty$. Then on $[0, \infty]^m$,*

$$\widetilde{\pi}_m^{(t)}\big(u_t\,,\,\cdot\,\big) := \mathsf{P}_{tu_t}\left[\left(\frac{X_1^{(t)}}{t}, \ldots, \frac{X_m^{(t)}}{t}\right) \in \cdot\,\right] \Rightarrow \mathsf{P}_u\big[(T_1^*,\, \ldots,\, T_m^*) \in \cdot\,\big] \qquad (t \to \infty).$$

Proof. Suppose $m \geq 2$ and $f \in \mathcal{C}[0,\infty]^m$, and assume first that $f \geq 0$. Then $f \in \mathcal{C}_K^+[0,\infty]^m$ as well, since the space is compact. Recall the notation of Theorem 3.2. Conditioning on $\tau(t)$, we can write

$$\begin{aligned}\widetilde{\pi}_m^{(t)}\big(u_t\,,\,\cdot\,\big)(f) &= \int_{(0,\infty]^m} f(\boldsymbol{x}_m)\,\widetilde{\pi}_m^{(t)}\big(u_t\,,\,d\boldsymbol{x}_m\big) + \sum_{k=1}^m \int_{(0,\infty]^{k-1}\times\{0\}^{m-k+1}} f(\boldsymbol{x}_m)\,\widetilde{\pi}_m^{(t)}\big(u_t\,,\,d\boldsymbol{x}_m\big)\\ &= \int_{(0,\infty]^m} f(\boldsymbol{x}_m)\,\widetilde{\pi}_m^{(t)}\big(u_t\,,\,d\boldsymbol{x}_m\big) + \sum_{k=1}^m \int_{(0,\infty]^{k-1}\times\{0\}} f(\boldsymbol{x}_k,\,0,\,\ldots,\,0)\,\widetilde{\pi}_k^{(t)}\big(u_t\,,\,d\boldsymbol{x}_k\big)\end{aligned}$$

by the Markov property. Since

$$\begin{aligned}\widetilde{\pi}_m^{(t)}\big(u_t\,,\,\cdot\cap(0,\infty]^m\big) &= \mathsf{P}_{tu_t}\big[t^{-1}\boldsymbol{X}_m^{(t)}\in\cdot\,,\ \tau(t)>m\big] = \mathsf{P}_{tu_t}\big[t^{-1}\boldsymbol{X}_m\in\cdot\cap(y(t),\infty]^m\big]\\ &= \widetilde{\mu}_{m+1}^{(t)}\big(u_t\,,\,\cdot\times[0,\infty]\big)\,,\end{aligned}$$

the first term becomes

$$\widetilde{\mu}_{m+1}^{(t)}\big(u_t\,,\,\cdot\big)(f) \longrightarrow \mu_{m+1}\big(u\,,\,\cdot\big)(f) = \int_{(0,\infty]^m} f(\boldsymbol{x}_m)\,\pi_m\big(u\,,\,d\boldsymbol{x}_m\big) = \int_{(0,\infty]^m} f(\boldsymbol{x}_m)\,\mathsf{P}_u\big[\boldsymbol{T}_m^*\in d\boldsymbol{x}_m\big]$$

as $t\to\infty$. Next, for any $A\subset[0,\infty]^k$ measurable, write $A_0=\{\boldsymbol{x}_{k-1}\,:\,(\boldsymbol{x}_{k-1}\,,\,0)\in A\}\subset[0,\infty]^{k-1}$, and observe that

$$\begin{aligned}\widetilde{\pi}_k^{(t)}\big(u_t\,,\,A\cap(0,\infty]^{k-1}\times\{0\}\big) &= \mathsf{P}_{tu_t}\big[t^{-1}\boldsymbol{X}_{k-1}^{(t)}\in A_0\cap(0,\infty]^{k-1}\,,\,X_k^{(t)}=0\big]\\ &= \mathsf{P}_{tu_t}\big[t^{-1}\boldsymbol{X}_{k-1}\in A_0\cap(y(t),\infty]^{k-1}\,,\ t^{-1}X_k\leq y(t)\big]\\ &= \widetilde{\mu}_k^{(t)}\big(u_t\,,\,A_0\times[0,\infty]\big)-\widetilde{\mu}_{k+1}^{(t)}\big(u_t\,,\,A_0\times[0,\infty]^2\big).\end{aligned}$$

Applying this reasoning to the terms in the summation yields

$$\begin{aligned}&\int_{[0,\infty]^k} f(\boldsymbol{x}_{k-1},\,0,\,\ldots,\,0)\,\widetilde{\mu}_k^{(t)}\big(u_t\,,\,d\boldsymbol{x}_k\big) - \int_{[0,\infty]^{k+1}} f(\boldsymbol{x}_{k-1},\,0,\,\ldots,\,0)\,\widetilde{\mu}_{k+1}^{(t)}\big(u_t\,,\,d\boldsymbol{x}_{k+1}\big)\\ &\qquad\longrightarrow \int_{[0,\infty]^k} f(\boldsymbol{x}_{k-1},\,0,\,\ldots,\,0)\,\mu_k\big(u\,,\,d\boldsymbol{x}_k\big) - \int_{[0,\infty]^{k+1}} f(\boldsymbol{x}_{k-1},\,0,\,\ldots,\,0)\,\mu_{k+1}\big(u\,,\,d\boldsymbol{x}_{k+1}\big)\\ &\qquad= \int_{(0,\infty]^{k-1}\times\{0\}} f(\boldsymbol{x}_k,\,0,\,\ldots,\,0)\,\pi_k\big(u\,,\,d\boldsymbol{x}_k\big) = \int_{(0,\infty]^{k-1}\times\{0\}^{m-k+1}} f(\boldsymbol{x}_m)\,\mathsf{P}_u\big[\boldsymbol{T}_m^*\in d\boldsymbol{x}_m\big].\end{aligned}$$

Combining these limits shows that $\mathsf{E}_{tu_t}f\big(t^{-1}\boldsymbol{X}_m^{(t)}\big)\longrightarrow\mathsf{E}_u f(\boldsymbol{T}_m^*)$, as $t\to\infty$. Finally, if f is not non-negative, then write $f=f_+-f_-$. Since each of f_+ and f_- is non-negative, bounded, and continuous, we can apply the above argument to each. □

4. The Regularity Condition

Previous work on the tail chain derives fdd convergence of $\boldsymbol{X}$ to $\boldsymbol{T}^*$ under a single assumption analogous to our domain of attraction condition (2.5). As we observed in Section 3.1, when $G(\{0\})=0$, fdd convergence of $\{t^{-1}\boldsymbol{X}\}$ follows directly, but when $G(\{0\})>0$, it was common to assume an additional technical condition which made (2.5) imply fdd convergence to $\boldsymbol{T}^*$ as well. This condition, which we refer to as the "regularity condition," is an asymptotic convergence assumption prescribing the boundary distribution to be $H=\epsilon_0$. We consider equivalences between different forms appearing in the literature, in terms of both kernels and update functions, and show that, under the regularity condition, the extremal behaviour of $\boldsymbol{X}$ is asymptotically the same as that of its extremal component $\boldsymbol{X}^{(t)}$.

In cases where $G(\{0\}) > 0$, Perfekt [18, 19] requires that

$$\lim_{\delta\downarrow 0}\ \limsup_{t\to\infty}\ \sup_{u\in[0,\delta]} K\big(tu\,,\,(t,\infty]\big) = 0, \tag{4.1}$$

while Segers [23] stipulates that the chosen update function corresponding to K must be of at most linear order in the initial state:

$$\limsup_{t\to\infty}\ \sup_{0\le y\le t} t^{-1}\psi(y,v) < \infty, \qquad (v\in B_0,\ \mathsf{P}[V\in B_0]=1). \tag{4.2}$$

Smith [24] used a variant of (4.1). We deem a formulation in terms of distributional convergence to be instructive in our context.

Definition. A Markov transition kernel $K \in D(G)$ satisfies the *regularity condition* if

$$K\big(tu_t\,,\,t\,\cdot\,\big) \Rightarrow \epsilon_0(\cdot) \tag{4.3}$$

on $[0,\infty]$ as $t\to\infty$ for any non-negative function $u_t = u(t) \to 0$.

Note that in (2.7) (p. 4), we had $u_t \to u > 0$. We interpret (4.3) as designating the boundary distribution H to be ϵ_0.

We now consider the relationships between (4.1), (4.2) and (4.3) , and propose an intuitive equivalent for update functions in canonical form.

Proposition 4.1. *Suppose $K \in D(G)$, and let $\psi(\cdot, V)$ be an update function corresponding to K such that*

$$t^{-1}\psi(t,v) \longrightarrow \xi(v) \tag{4.4}$$

whenever $v\in B$ for which $\mathsf{P}[V\in B]=1$, and $\xi\circ V \sim G$. Then:

(a) *Condition* (4.1) *is necessary and sufficient for K to satisfy the regularity condition* (4.3)*.*
(b) *Condition* (4.2) *is sufficient for K to satisfy the regularity condition* (4.3)*.*
(c) *If ψ is in canonical form, i.e.*

$$\psi(y,(Z,W)) = Zy + \phi(y,W),$$

then ψ satisfies (4.2) *if and only if $\phi(\cdot,w)$ is bounded on any neighbourhood of* 0 *for each $w\in C$, a set for which $\mathsf{P}[W\in C]=1$.*

Proof. (a) Assume (4.1), and suppose $u_t \to 0$. We show $K(tu_t\,,\,t(x,\infty]) \to 0$ for any $x>0$. Write

$$\omega(t,\delta) = \sup_{u\in[0,\delta]} K\big(tu\,,\,(t,\infty]\big).$$

Let $\epsilon > 0$ be given, and choose δ small enough that $\limsup_{t\to\infty}\omega(t,\delta) < \epsilon/2$. Then for t large enough that $u_t < \delta x$, we have

$$K\big(tu_t\,,\,t(x,\infty]\big) \le \sup_{u\in[0,\delta x]} K\big(tu\,,\,t(x,\infty]\big) = \omega(tx,\delta) < \limsup_{t\to\infty}\ \omega(t,\delta) + \epsilon/2$$

for t large enough. Our choice of δ implies that $K(tu_t\,,\,t(x,\infty]) < \epsilon$.

Conversely, assume that K satisfies (4.3) but that (4.1) fails. Choose $\epsilon > 0$ and a sequence $\delta_n \downarrow 0$ such that $\limsup_{t\to\infty}\omega(t,\delta_n) \ge \epsilon$ for $n = 1,2,\dots$. Then for each n we can find a sequence $t^n_k \to\infty$ as $k\to\infty$ such that $\omega(t^n_k,\delta_n)\ge\epsilon$ for each k. Diagonalize to find $k_1 < k_2 < \cdots$ such that $s_n = t^n_{k_n} \to \infty$ and $\omega(s_n,\delta_n)\ge\epsilon$ for all n. Finally, for $n=1,2,\dots$ choose $u_n\in[0,\delta_n]$ such that

$$K\big(s_n u_n\,,\,(s_n,\infty]\big) > \omega(s_n,\delta_n) - \epsilon/2,$$

and put $u(t) = \sum_n u_n\,\mathbf{1}_{[s_n,s_{n+1})}(t)$. Clearly $u(t)\to 0$, but $K(s_n u(s_n)\,,\,(s_n,\infty]) \ge \epsilon/2$ for all n, contradicting (4.3).

(b) Write $M(v) = \limsup_t \sup_{0\le y\le t} t^{-1}\psi(y,v)$. Since

$$\sup_{0\le y\le t} t^{-1}\psi(y,v) = \sup_{0\le y\le \delta} \frac{\psi(t\delta^{-1}y,v)}{t\delta^{-1}}\,\delta^{-1}$$

for $\delta > 0$, we have

$$\limsup_{t\to\infty} \sup_{0\le y\le \delta} t^{-1}\psi(ty,v) = \delta M(v).$$

Now, suppose $u_t \to 0$. Given any $\delta > 0$ we have

$$t^{-1}\psi(tu_t,v) \le \sup_{0\le y\le \delta} t^{-1}\psi(ty,v)$$

provided t is large enough, so $\limsup_t t^{-1}\psi(tu_t,v) \le \delta M(v)$. Consequently, $\limsup_t t^{-1}\psi(tu_t,v) = 0$ for every v such that $M(v) < \infty$. Under (4.2), this means that $\mathsf{P}\big[t^{-1}\psi(tu_t,V) \to 0\big] = 1$, implying (4.3).

(c) Suppose first that $\chi_w(a) = \sup_{0\le y\le a} \phi(y,w) < \infty$ for all $a > 0$, whenever $w \in C$. Fixing $w \in C$ and $z \ge 0$, note that

$$\sup_{0\le y\le t} t^{-1}\psi(y,(z,w)) \le z + \sup_{0\le y\le t} t^{-1}\phi(y,w),$$

and observe for any $a > 0$ that

$$\sup_{0\le y\le t} t^{-1}\phi(y,w) \le \Big(\sup_{0\le y\le a} t^{-1}\phi(y,w)\Big) \vee \Big(\sup_{a\le y\le t} y^{-1}\phi(y,w)\Big) \le t^{-1}\chi_w(a) \vee \Big(\sup_{a\le y} y^{-1}\phi(y,w)\Big).$$

Choosing a large enough that $\sup_{a\le y} y^{-1}\phi(y,w) \le 1$, say, it follows that

$$\limsup_{t\to\infty} \sup_{0\le y\le t} t^{-1}\psi(y,(z,w)) \le z+1,$$

so $v = (z,w) \in B_0$. Therefore $\mathsf{P}[(Z,W) \in B_0] \ge \mathsf{P}[Z \ge 0,\ W \in C] = 1$.

Conversely, suppose there is a set D with $\mathsf{P}[W \in D] > 0$ such that $w \in D$ implies $\chi_w(a) = \infty$ for some $0 < a < \infty$. Since $\sup_{0\le y\le t} t^{-1}\psi(y,(z,w)) \ge t^{-1}\chi_w(t)$, we have $[0,\infty)\times D \subset B_0^{\mathrm{c}}$, contradicting (4.2). $\square$

The exclusion of necessity from part (b) results from the fact that a kernel K does not uniquely specify an update function ψ. Even when K satisfies the regularity condition (4.3), it may be possible to choose a nasty update function ψ which satisfies (4.4), but not (4.2). However, in such cases there may exist a different update function ψ' corresponding to K which does satisfy (4.2).

Here is an example of such a situation. We exhibit an update function ψ for which (i) (4.4) holds; (ii) (4.2) fails because condition (c) in Proposition 4.1 fails; but yet (iii) the corresponding kernel satisfies the regularity condition (4.3). Furthermore, we present a different choice of update function corresponding to the same kernel which satisfies (4.2). Define $\psi(y, V = (Z,W)) = Zy + \phi(y,W)$, where

$$\phi(y,w) = \sum_{k=1}^{\infty} k\cdot \mathbf{1}_{\{yw=1/k\}}$$

and $W \sim U(0,1)$. (i) Since $\phi(t,w) = 0$ for $t > 1/w$, it is clear that ψ satisfies (4.4) with $\xi = Z$. (ii) Observe that for any $w \in (0,1)$, $\phi(\cdot,w)$ is unbounded on the interval $[0,1]$. Therefore, by part (c) of Proposition 4.1, (4.2) cannot hold for ψ. (iii) However, the corresponding kernel does satisfy the regularity condition (4.3). Suppose $u_t \to 0$ and $a > 0$ is arbitrarily large. Write

$$\mathsf{P}\big[t^{-1}\psi(tu_t,(Z,W)) > x\big] = \mathsf{P}\big[Zu_t + t^{-1}\phi(tu_t,W) > x\big] \le \mathsf{P}\big[t^{-1}\phi(tu_t,W) > x'\big] + \mathsf{P}[Z > a],$$

choosing $0 < x' < x - au_t$. Since for any t, $\{w : \phi(tu_t, w) > tx'\} \subset \{(tu_t k)^{-1} : k = 1, 2, \dots\}$, a set of measure 0 with respect to $\mathsf{P}[W \in \cdot\,]$, (4.3) follows by letting $a \to \infty$. On the other hand, the update function $\psi'(y, Z) = Zy$ does satisfy (4.2), and for any y,

$$\mathsf{P}\big[\psi'(y, Z) \neq \psi(y, (Z, W))\big] = \mathsf{P}\big[W \in \{(yk)^{-1} : k = 1, 2, \dots\}\big] = 0,$$

so ψ' does indeed correspond to K.

The regularity condition (4.3) restricts attention to Markov chains for which the probability of returning to an extreme state in the next m steps after falling below the extremal boundary is asymptotically negligible. For such chains, as well as those for which $y(t) \equiv 0$ is an extremal boundary for K, $\boldsymbol{X}$ has the same asymptotic behaviour as its extremal component, as described next.

Theorem 4.1. *Suppose $\boldsymbol{X} \sim K$ with $K \in D(G)$, and let ρ be a metric on $\mathbb{R}^m$. If $y(t) \equiv 0$ is an extremal boundary for K, or if K satisfies the regularity condition* (4.3)*, then for any $\epsilon > 0$ we have*

$$\mathsf{P}_{tu_t}\Big[\rho\Big(\frac{\boldsymbol{X}_m^{(t)}}{t}\,,\,\frac{\boldsymbol{X}_m}{t}\Big) > \epsilon\Big] \longrightarrow 0 \qquad (t \to \infty,\ u_t \to u > 0). \tag{4.5}$$

Consequently,

$$\mathsf{P}_{tu_t}\Big[\Big(\frac{X_1}{t}, \dots, \frac{X_m}{t}\Big) \in \cdot\,\Big] \Rightarrow \mathsf{P}_u\big[(T_1^*, \dots, T_m^*) \in \cdot\,\big] \qquad (t \to \infty,\ u_t \to u > 0). \tag{4.6}$$

First let us extend the regularity condition to higher-order transition kernels.

Lemma 4.1. *If K satisfies* (4.3)*, then so do the m-step transition kernels K^m.*

Proof. This is established by induction. Let $u_t \to 0$ and $f \in \mathcal{C}[0, \infty]$. For $m \geq 2$, we have

$$K^m\big(tu_t\,,\,\cdot\big)(f) = \int_{[0,\infty]} K^{m-1}\big(tu_t\,,\,tdv\big) \int_{[0,\infty]} K\big(tv\,,\,tdx\big)\, f(x).$$

Assume that $K^{m-1}(tu_t\,,\,t\,\cdot) \Rightarrow \epsilon_0$; (4.3) implies that $\int K(tv_t\,,\,tdx)\, f(x) \to f(0)$ whenever $v_t \to 0$. Therefore, by Lemma 8.2 (a) (p. 21), we conclude that

$$K^m\big(tu_t\,,\,\cdot\big)(f) \longrightarrow f(0) = \epsilon_0(f). \qquad \square$$

Proof of Theorem 4.1. Suppose $\epsilon > 0$ and $u_t \to u > 0$. Write

$$\mathsf{P}_{tu_t}\big[\rho(t^{-1}\boldsymbol{X}_m^{(t)}\,,\,t^{-1}\boldsymbol{X}_m) > \epsilon\big] = \sum_{k=1}^{m} \mathsf{P}_{tu_t}\big[\rho(t^{-1}\boldsymbol{X}_m^{(t)}\,,\,t^{-1}\boldsymbol{X}_m) > \epsilon\,,\ \tau(t) = k\big].$$

Since $X_j = X_j^{(t)}$ while $j < \tau(t)$, for the k-th summand to converge to 0, it is sufficient that

$$\mathsf{P}_{tu_t}\big[|X_j^{(t)}/t - X_j/t| > \delta\,,\ \tau(t) = k\big] = \mathsf{P}_{tu_t}\big[X_j/t > \delta\,,\ \tau(t) = k\big] \longrightarrow 0$$

for $j = k, \dots, m$ and any $\delta > 0$. If $j = k$, we have

$$\mathsf{P}_{tu_t}\big[X_j/t > \delta\,,\ \tau(t) = k\big] \leq \mathsf{P}_{tu_t}\big[X_k/t > \delta\,,\ X_k/t \leq y(t)\big] = 0$$

for large t. For $j > k$, recalling the notation of Theorem 3.2,

$$\begin{aligned}
\mathsf{P}_{tu_t}\big[X_j/t > \delta\,,\ \tau(t) = k\big] &= \int_{\mathbb{E}_k'(t)} \mathbf{1}_{[0,y(t)]}(x_k)\, \mathsf{P}_{tu_t}\big[X_j/t > \delta \big| \boldsymbol{X}_k/t = \boldsymbol{x}_k\big]\ \mathsf{P}_{tu_t}\big[\boldsymbol{X}_k/t \in d\boldsymbol{x}_k\big] \\
&= \int_{[0,\infty]^k} \mathsf{P}_{tx_k}\big[X_{j-k} > t\delta\big]\, \mathbf{1}_{[0,y(t)]}(x_k)\ \widetilde{\mu}_k^{(t)}\big(u_t\,,\,d\boldsymbol{x}_k\big)
\end{aligned}$$

using the Markov property. We claim that this intergral $\to 0$ as $t \to \infty$. If $y(t) \equiv 0$, this follows directly. Otherwise, recall that $\widetilde{\mu}_k^{(t)}(u_t\,,\,\cdot) \xrightarrow{v} \mu_k(u\,,\,\cdot)$, and consider $h_t(\boldsymbol{x}_k) = \mathsf{P}_{tx_k}[X_{j-k} > t\delta]\,\mathbf{1}_{[0,y(t)]}(x_k)$. Suppose $\boldsymbol{x}^{(t)} \to \boldsymbol{x} \in [0,\infty]^k$. If $x_k > 0$, then $h_t(\boldsymbol{x}^{(t)}) = 0$ for large t because $y(t) \to 0$. Otherwise, if $x_k = 0$, we have $h_t(\boldsymbol{x}^{(t)}) \to 0$ since Lemma 4.1 implies that $\mathsf{P}_{tx_k^{(t)}}[X_{j-k} > t\delta] \to 0$ as $t \to \infty$. Lemma 8.2 (b) establishes (4.5); (4.6) follows by Slutsky's theorem. $\square$

Therefore, $\boldsymbol{X}$ converges to $\boldsymbol{T}^*$ in fdds under (a) $G(\{0\}) = 0$, (b) $G(\{0\}) > 0$ combined with (4.3), or (c) $G(\{0\}) > 0$ combined with the extremal boundary $y(t) \equiv 0$. In either case, we will be able to replace the extremal component $\boldsymbol{X}^{(t)}$ with the complete chain $\boldsymbol{X}$ in the results of Sections 5.1 and 5.2. However, that $y(t) \equiv 0$ is an extremal boundary, and consequently that (4.6) holds, does not imply the regularity condition to hold, regardless of $G(\{0\})$; in particular, a kernel for which $G(\{0\}) = 0$ need not satisfy (4.3). This is illustrated in Example 6.3.

5. Convergence of the Unconditional FDDs

5.1. **Effect of a Regularly Varying Initial Distribution.** So far our convergence results required that the initial state become large, and the only distributional assumption was that the transition kernel K determining $\boldsymbol{X}$ be attracted to some distribution G. To obtain a result for the unconditional distribution of $(X_0, \ldots, X_m)$, we require an additional assumption about how likely the initial observation X_0 is to be large. Using Lemma 8.4, the results of the previous sections extend to multivariate regular variation on the cone $\mathbb{E}_m = (0,\infty] \times [0,\infty]^m$ when the distribution of X_0 has a regularly varying tail. This cone is smaller than the cone $[0,\infty]^{m+1} \backslash \{\mathbf{0}\}$ traditionally employed in extreme value theory, because the kernel domain of attraction condition (2.5) is uninformative when the initial state is not extreme. This is analogous to the setting of the Conditional Extreme Value Model considered in [8, 13].

Proposition 5.1. *Assume $\boldsymbol{X} \sim K$ with $K \in D(G)$, and $X_0 \sim H$, where H is a distribution on $[0,\infty)$ with a regularly varying tail. This means that as $t \to \infty$, for some scaling function $b(t) \to \infty$,*

$$tH\big(b(t)\,\cdot\,\big) \xrightarrow{v} \nu_\alpha(\cdot) \quad \text{in } \mathbb{M}_+(0,\infty],$$

where $\nu_\alpha(x,\infty] = x^{-\alpha}$ and $\alpha > 0$. Define the measure ν^* on $\mathbb{E}_m = (0,\infty] \times [0,\infty]^m$ by

$$\nu^*\big(dx_0, d\boldsymbol{x}_m\big) = \nu_\alpha(dx_0)\,\mathsf{P}_{x_0}\big[(T_1^*, \ldots, T_m^*) \in d\boldsymbol{x}_m\big]. \tag{5.1}$$

Then, for $m = 1, 2, \ldots$, the following convergences take place as $t \to \infty$:

(a) *In $\mathbb{M}_+((0,\infty]^m \times [0,\infty])$,*

$$t\,\mathsf{P}\big[b(t)^{-1}(X_0\,, X_1\,, \ldots, X_m) \in \cdot \cap (0,\infty]^m \times [0,\infty]\big] \xrightarrow{v} \nu^*\big(\,\cdot \cap (0,\infty]^m \times [0,\infty]\big).$$

(b) *In $\mathbb{M}_+(\mathbb{E}_m)$,*

$$t\,\mathsf{P}\big[b(t)^{-1}(X_0^{(b(t))}\,, X_1^{(b(t))}\,, \ldots, X_m^{(b(t))}) \in \cdot\,\big] \xrightarrow{v} \nu^*(\cdot).$$

(c) *If either $G(\{0\}) = 0$, $y(t) \equiv 0$ is an extremal boundary, or K satisfies the regularity condition (4.3), then in $\mathbb{M}_+(\mathbb{E}_m)$,*

$$t\,\mathsf{P}\big[b(t)^{-1}(X_0\,, X_1\,, \ldots, X_m) \in \cdot\,\big] \xrightarrow{v} \nu^*(\cdot).$$

(d) *In $\mathbb{M}_+(0,\infty]$,*

$$t\,\mathsf{P}\big[X_0/b(t) \in dx_0\,,\ \tau(b(t)) \geq m\big] \xrightarrow{v} \big(1 - G(\{0\})\big)^{m-1} \cdot \nu_\alpha(dx_0).$$

Remark. These convergence statements may be reformulated equivalently as, say,

$$\mathsf{P}\big[b(t)^{-1}(X_0, X_1, \ldots, X_m) \in \cdot \;\big|\; X_0 > b(t)\big] \Rightarrow \mathsf{P}\big[(T_0^*, T_1^*, \ldots, T_m^*) \in \cdot\,\big],$$

where $T_0^* \sim \text{Pareto}(\alpha)$. This is the form considered by Segers [23].

Proof. Apply Lemma 8.4 (p. 22) to the results of Theorems 3.1, 3.3 and 4.1, and (3.15). $\square$

In the case $m = 1$, $\mathbb{E}_1$ is a rotated version of $\mathbb{E}_\sqcap$ used in the conditional extreme value model in [8, 9] and the limit can be expressed as

$$\nu^*\big((x_0, \infty] \times [0, x_1]\big) = \int_{x_0}^{\infty} \nu_\alpha(du) P[\xi \le x_1/u] = x_0^{-\alpha}\, \mathsf{P}\big[\xi \le x_1/x_0\big] - x_1^{-\alpha}\, \mathsf{E}\xi^\alpha \mathbf{1}_{\{\xi \le x_1/x_0\}}$$

for $(x_0, x_1) \in (0, \infty] \times [0, \infty]$, where $\xi \sim G$ (with $\mathsf{E}\xi^\alpha \le \infty$). Since

$$\nu^*\big((x_0, \infty] \times \{0\}\big) = x_0^{-\alpha}\, \mathsf{P}[\xi = 0] \qquad \text{and} \qquad \nu^*\big((0, \infty] \times (x_1, \infty]\big) = x_1^{-\alpha}\, \mathsf{E}\xi^\alpha,$$

sets on the x_0-axis incur mass proportional to $G(\{0\})$, and sets bounded away from this axis are weighted accordng to $\mathsf{E}\xi^\alpha$. A consequence of the second observation is that

$$\liminf_{t \to \infty} t\, \mathsf{P}\big[X_1/b(t) > x\big] \ge \mathsf{E}\xi^\alpha \cdot x^{-\alpha}.$$

Thus, knowledge concerning the tail behaviour of X_1 imposes a restriction on the distributions G to which K can be attracted via the α-th moment. For example, if $t\, \mathsf{P}[X_1/b(t) \in \cdot\,] \xrightarrow{v} \nu_\alpha$, then we must have $\mathsf{E}\xi^\alpha \le 1$; this property will be examined further in the next section and appears in various forms in Segers [23] and Basrak and Segers [1], in the stationary setting.

5.2. **Joint Tail Convergence.** What additional assumptions are necessary for convergences (b) and (c) of the previous result to take place on the larger cone $\mathbb{E}_m^* = [0, \infty]^{m+1} \setminus \{\mathbf{0}\}$? This was considered by Segers [1, 23] for stationary Markov chains. In (b), the dependence on the extremal threshold and hence on t means we are in the context of a triangular array and not, strictly speaking, in the setting of joint regular variation. However, the result is still useful, for example, to derive a point process convergence via the Poisson transform [21, p. 183].

As a first step, we characterize convergence on the larger cone by decomposing it into smaller, more familiar cones. This is similar to Theorem 6.1 in [23] and one of the implications of Theorem 2.1 in [1]. As a convention in what follows, set $[0, \infty]^0 \times A = A$. Also, recall the notation $\mathbb{E}_m = (0, \infty] \times [0, \infty]^m$.

Proposition 5.2. *Suppose $\boldsymbol{Y}_t = (Y_{t,0}\,, Y_{t,1}\,, \ldots, Y_{t,m})$ is a random vector on $[0, \infty]^{m+1}$ for each $t > 0$. Then there exists a non-null Radon measure μ^* on $\mathbb{E}_m^* = [0, \infty]^{m+1} \setminus \{\mathbf{0}\}$ such that*

$$t\, \mathsf{P}\big[(Y_{t,0}\,, Y_{t,1}\,, \ldots, Y_{t,m}) \in \cdot\,\big] \xrightarrow{v} \mu^*(\cdot) \quad \text{in } \mathbb{M}_+(\mathbb{E}_m^*) \qquad (t \to \infty) \tag{5.2}$$

if and only if for $j = 0, \ldots, m$ there exist Radon measures μ_j on $\mathbb{E}_j = (0, \infty] \times [0, \infty]^j$, not all null, such that

$$t\, \mathsf{P}\big[(Y_{t,j}\,, \ldots, Y_{t,m}) \in \cdot\,\big] \xrightarrow{v} \mu_{m-j}(\cdot) \quad \text{in } \mathbb{M}_+(\mathbb{E}_{m-j}). \tag{5.3}$$

The relation between the limit measures is the following:

$$\mu_{m-j}(\cdot) = \mu^*\big([0, \infty]^j \times \cdot\,\big) \quad \text{on } \mathbb{E}_{m-j}$$

for $j = 0, \ldots, m$, and

$$\mu^*\big([\mathbf{0}, \boldsymbol{x}]^c\big) = \sum_{j=0}^{m} \mu_{m-j}\big((x_j, \infty] \times [0, x_{j+1}] \times \cdots \times [0, x_m]\big) \quad \text{for } \boldsymbol{x} \in \mathbb{E}_m^*.$$

Furthermore, given $j \in \{0, \ldots, m-1\}$, if $A \subset [0, \infty]^{m-j} \setminus \{0\}^{m-j}$ is relatively compact, then $\mu_{m-j}((0, \infty] \times A) < \infty$.

Proof. Assume first that (5.2) holds. Fixing $j \in \{0, \ldots, m\}$, define $\mu_{m-j}(\cdot) := \mu^*([0,\infty]^j \times \cdot)$ (i.e. $\mu_m = \mu^*$). Let $A \subset \mathbb{E}_{m-j}$ be relatively compact with $\mu_{m-j}(\partial A) = 0$. Then $A^* = [0,\infty]^j \times A$ is relatively compact in $\mathbb{E}_m^*$, and $\partial_{\mathbb{E}_m^*} A^* = [0,\infty]^j \times \partial_{\mathbb{E}_{m-j}} A$, so $\mu^*(\partial_{\mathbb{E}_m^*} A^*) = \mu_{m-j}(\partial A) = 0$. Therefore,

$$t\,\mathsf{P}\big[(Y_{t,j}\,,\,\ldots,\,Y_{t,m}) \in A\big] = t\,\mathsf{P}\big[(Y_{t,0}\,,\,\ldots,\,Y_{t,m}) \in A^*\big] \longrightarrow \mu^*(A^*) = \mu_{m-j}(A),$$

establishing (5.3).

Conversely, suppose we have (5.3) for $j = 0, \ldots, m$. For $\boldsymbol{x} \in (0,\infty]^{m+1}$, define

$$h(\boldsymbol{x}) = \sum_{j=0}^{m} \mu_{m-j}\big((x_j,\infty] \times [0,x_{j+1}] \times \cdots \times [0,x_m]\big).$$

Decompose $[\mathbf{0},\boldsymbol{x}]^{\mathrm{c}}$ as a disjoint union

$$[\mathbf{0},\boldsymbol{x}]^{\mathrm{c}} = \bigcup_{j=0}^{m} [0,\infty]^j \times (x_j,\infty] \times [0,x_{j+1}] \times \cdots \times [0,x_m]\,, \tag{5.4}$$

and observe that at points of continuity of the limit,

$$t\,\mathsf{P}\big[\boldsymbol{Y}_t \in [\mathbf{0},\boldsymbol{x}]^{\mathrm{c}}\,\big] = \sum_{j=0}^{m} t\,\mathsf{P}\big[(Y_{t,j}\,,\,\ldots,\,Y_{t,m}) \in (x_j,\infty] \times [0,x_{j+1}] \times \cdots \times [0,x_m]\,\big] \longrightarrow h(\boldsymbol{x}). \tag{5.5}$$

Hence, (5.2) holds with the limit measure μ^* defined by $\mu^*\big([\mathbf{0},\boldsymbol{x}]^{\mathrm{c}}\big) = h(\boldsymbol{x})$. Indeed, given $f \in \mathcal{C}_K^+(\mathbb{E}_m^*)$ we can find $\delta > 0$ such that $\boldsymbol{x}_\delta = (\delta, \ldots, \delta)$ is a continuity point of h and f is supported on $[\mathbf{0},\boldsymbol{x}_\delta]^{\mathrm{c}}$. Therefore,

$$t\,\mathsf{E}f(\boldsymbol{Y}_t) \le \sup_{\boldsymbol{x}\in\mathbb{E}_m^*} f(\boldsymbol{x}) \cdot \sup_{t>0} t\,\mathsf{P}\big[\boldsymbol{Y}_t \in [\mathbf{0},\boldsymbol{x}_\delta]^{\mathrm{c}}\big] < \infty,$$

implying that the set $\big\{t\,\mathsf{P}[\boldsymbol{Y}_t \in \cdot\,];\ t > 0\big\}$ is relatively compact in $\mathbb{M}_+(\mathbb{E}_m^*)$. Furthermore, if $t_k\,\mathsf{P}[\boldsymbol{Y}_{t_k} \in \cdot\,] \to \mu$ and $s_k\,\mathsf{P}[\boldsymbol{Y}_{s_k} \in \cdot\,] \to \mu'$ as $k \to \infty$, then $\mu = \mu' = \mu^*$ on sets $[\mathbf{0},\boldsymbol{x}]^{\mathrm{c}}$ which are continuity sets of μ^* by (5.5). This extends to measurable rectangles in $\mathbb{E}_m^*$ bounded away from $\mathbf{0}$ whose vertices are continuity points of h, leading us to the conclusion that $\mu = \mu' = \mu^*$ on $\mathbb{E}_m^*$.

Moreover, since we can decompose $[\mathbf{0},\boldsymbol{x}]^{\mathrm{c}}$ for any $\boldsymbol{x} \in \mathbb{E}_m^*$ as in (5.4), it is clear that μ^* is non-null iff not all of the μ_j are null.

Finally, for $1 \le j \le m-1$, if $A \subset [0,\infty]^{m-j}\backslash\{0\}^{m-j}$ is relatively compact, then it is contained in $[(0,\ldots,0),(x_{j+1},\ldots,x_m)]^{\mathrm{c}}$ for some $(x_{j+1},\ldots,x_m) \in (0,\infty]^{m-j}$. Applying (5.4) once again, we find that

$$\begin{aligned}
\mu_{m-j}\big((0,\infty] \times A\big) &= \mu^*\big([0,\infty]^j \times (0,\infty] \times A\big) \\
&\le \sum_{k=j+1}^{m} \mu^*\big([0,\infty]^{j+1} \times [0,\infty]^{k-j-1} \times (x_k,\infty] \times [0,x_{k+1}] \times \cdots \times [0,x_m]\big) \\
&= \sum_{k=j+1}^{m} \mu_{m-k}\big((x_k,\infty] \times [0,x_{k+1}] \times \cdots \times [0,x_m]\big) < \infty.
\end{aligned}$$

□

Consequently, the extension of the convergences in Proposition 5.1 to the larger cone $\mathbb{E}_m^*$ follows from regular variation of the marginal tails.

Theorem 5.1. *Suppose $\boldsymbol{X} \sim K \in D(G)$, and let $b(t) \to \infty$ be a scaling function and $\alpha > 0$. Then*

$$t\,\mathsf{P}\big[b(t)^{-1}(X_0^{(b(t))},\,X_1^{(b(t))},\,\ldots,\,X_m^{(b(t))}) \in \cdot\,\big] \xrightarrow{v} \mu^*(\cdot) \qquad \textit{in } \mathbb{M}_+(\mathbb{E}_m^*) \qquad (t \to \infty), \tag{5.6}$$

where

$$\mu^*\big|_{\mathbb{E}_m}(dx_0, d\boldsymbol{x}) = \nu_\alpha(dx_0)\,\mathsf{P}_{x_0}\big[(T_1^*,\,\dots,\,T_m^*)\in d\boldsymbol{x}_m\big] = \nu^*(dx_0, d\boldsymbol{x}),$$

if and only if

$$t\,\mathsf{P}\big[X_j^{(b(t))}/b(t)\in\cdot\,\big] \xrightarrow{v} c_j\,\nu_\alpha(\cdot) \tag{5.7}$$

in $\mathbb{M}_+(0,\infty]$*, with* $c_0 = 1$ *and* $(\mathsf{E}\xi^{\,\alpha})^j \le c_j < \infty$ *for* $j = 1,\dots,m$*.*

Proof. Assume first that (5.6) holds. It follows that

$$t\,\mathsf{P}\big[X_0 > b(t)x\big] \longrightarrow \nu^*((x,\infty]\times[0,\infty]^m) = x^{-\alpha}$$

for $x>0$. Hence, $b(t)\in\mathrm{RV}_{1/\alpha}$, so by (5.6) again, we have for $j\ge 1$

$$t\,\mathsf{P}\big[X_j^{(b(t))} > b(t)x\big] \longrightarrow \mu^*([0,\infty]^j\times(x,\infty]\times[0,\infty]^{m-j}) = c_j\,x^{-\alpha},$$

and

$$\begin{aligned} c_j \ge \mu^*((0,\infty]\times[0,\infty]^{j-1}\times(1,\infty]\times[0,\infty]^{m-j}) &= \int_{(0,\infty]}\nu_\alpha(du)\;\mathsf{P}\big[\xi_1\cdots\xi_j > u^{-1}\big]\\ &= \mathsf{E}(\xi_1\cdots\xi_j)^\alpha = (\mathsf{E}\xi^\alpha)^j. \end{aligned}$$

Conversely, suppose that (5.7) holds for $j=0,\dots,m$. Lemma 8.4 implies that in $\mathbb{M}_+(\mathbb{E}_{m-j})$,

$$\begin{aligned} t\,\mathsf{P}\big[b(t)^{-1}(X_j^{(b(t))},\dots,X_m^{(b(t))})\in(dx_0,d\boldsymbol{x})\,\big] &\xrightarrow{v} c_j\,\nu_\alpha(dx_0)\,\mathsf{P}_{x_0}\big[(T_1^*,\dots,T_{m-j}^*)\in d\boldsymbol{x}\big]\\ &=: \mu_{m-j}\big((dx_0,d\boldsymbol{x})\big) \end{aligned}$$

by the Markov property, and Proposition 5.2 yields (5.6), with $\mu^*|_{\mathbb{E}_m}(\cdot) = \mu_m(\cdot) = \nu^*(\cdot)$. □

At the end of Section 4, cases were outlined in which we could replace $X_j^{(b(t))}$ by X_j. Theorem 5.1 is most striking for these since it shows that for a Markov chain whose kernel is in a domain of attraction, to obtain joint regular variation of the fdds it is enough to know that the marginal tails are regularly varying. In particular, if $\boldsymbol{X}$ has a regularly varying stationary distribution then the fdds are jointly regularly varying. This result was presented by Segers [23], and Basrak and Segers [1] showed that for a general stationary process, joint regular variation of fdds is equivalent to the existence of a "tail process" which reduces to the tail chain in the case of Markov chains. However, what Proposition 5.1 emphasizes is that it is the marginal tail behaviour alone, rather than stationarity, which provides the link with joint regular variation.

Theorem 5.1 also extends the observation made in Section 5.1 that knowledge of the marginal tail behaviour for a Markov chain whose kernel is in a domain of attraction constrains the class of possible limit distributions G via its moments. If a particular choice of regularly varying initial distribution leads to $t\,\mathsf{P}[X_j > b(t)\,\cdot\,] \xrightarrow{v} a_j\nu_\alpha(\cdot)$, then we have $\mathsf{E}\xi^\alpha \le a_j^{1/j}$. In particular, if $\boldsymbol{X}$ admits a stationary distribution whose tail is $\mathrm{RV}_{-\alpha}$, then $\mathsf{E}\xi^\alpha\le 1$.

6. Examples

Our first example illustrates the main results.

***Example* 6.1.** Let $V=(Z,\,W)$ be any random vector on $[0,\infty)\times\mathbb{R}$. Consider the update function $\psi(y,V) = (Zy+W)_+$ and its canonical form

$$\psi(y,V) = Zy + \phi(y,W) = Zy + \big(W\,\mathbf{1}_{\{W>-Zy\}} - Zy\,\mathbf{1}_{\{W\le -Zy\}}\big).$$

For $y>0$, the transition kernel has the form $K(y\,,\,(x,\infty)) = \mathsf{P}\,[Zy+W>x]$. Since $t^{-1}\psi(t,V) = (Z+t^{-1}W)_+\to Z$ a.s., we have $K\in D(G)$ with $G=\mathsf{P}[Z\in\cdot\,]$. Furthermore, using Proposition 3.1, the function $\gamma(t)\equiv\sqrt{t}$ is of larger order than $\phi(t,w)$, so $y(t) = 1/\sqrt{t}$ is an extremal boundary. Since

$\phi(\cdot, w)$ is bounded on neighbourhoods of 0, Proposition 4.1 (c) implies K satisfies the regularity condition (4.3). Consequently, from Theorem 4.1, we obtain fdd convergence of $t^{-1}\boldsymbol{X}$ to $\boldsymbol{T}^*$ as in (4.6). $\square$

If K does not satisfy the regularity condition (4.3), Theorem 4.1 may fail to hold and starting from tu, $t^{-1}\boldsymbol{X}$ may fail to converge to $\boldsymbol{T}^*$ started from u.

***Example* 6.2.** Let $V = (Z,\, W,\, W')$ be any non-degenerate random vector on $[0,\infty)^3$, and consider the Markov chain determined by the update function

$$\psi(y, V) = Zy + W\, y^{-1}\, \mathbf{1}_{\{y>0\}} + W'\, \mathbf{1}_{\{y=0\}}.$$

For $y > 0$, the transition kernel is $K(y\,,\, (x,\infty)) = \mathsf{P}[Zy + Wy^{-1} > x]$ and since $t^{-1}\psi(t, V) = Z + Wt^{-2} \to Z$ a.s., we have $K \in D(G)$ with $G = \mathsf{P}[Z \in \cdot\,]$. Furthermore, using Proposition 3.1, the function $\gamma(t) \equiv 1$ is of larger order than $\phi(t, w)$, so $y(t) = 1/t$ is an extremal boundary.

However, note that $\phi(y, (W, W')) = Wy^{-1}\mathbf{1}_{\{y>0\}} + W'\mathbf{1}_{\{y=0\}}$ is unbounded near 0, implying that Segers' boundedness condition (4.2) does not hold. In fact, our form of the regularity condition (4.3) fails for K. Indeed,

$$K\big(tu_t\,,\, t(x,\infty)\big) = \mathsf{P}[Ztu_t + W/(tu_t) > tx] = \mathsf{P}[Zu_t + W/(t^2u_t) > x].$$

Choosing $u_t = t^{-2}$ yields $K(tu_t\,,\, t(x,\infty)) \to \mathsf{P}[W > x]$. For appropriate x, this shows (4.3) fails.

Not only does (4.3) fail but so does Theorem 4.1, since the asymptotic behaviour of $\boldsymbol{X}$ is not the same as that of $\boldsymbol{X}^{(t)}$. We show directly that the conditional fdds of $t^{-1}\boldsymbol{X}$ fail to converge to those of $\boldsymbol{T}^*$. The idea is that if $X_k < y(t) = t^{-1}$, there is a positive probability that $X_{k+1} > t$. We illustrate this for $m = 2$. Take $f \in \mathcal{C}[0,\infty]^2$ and $u > 0$. Observe if $X_0 = tu > 0$, from the definition of ψ, $X_1 = Z_1tu + W_1/(tu)$ and $X_2 = Z_2X_1 + (W_2/X_1)\mathbf{1}_{\{X_1>0\}} + W'\mathbf{1}_{\{X_1=0\}}$. Furthermore, on $\{Z_1 > 0\}$, we have $X_1 > 0$ and $X_2 = Z_2X_1 + W_2/X_1$. On $\{Z_1 = 0, W_1 > 0\}$, $X_1 > 0$ and $X_2 = Z_2X_1 + W_2/X_1$. On $\{Z_1 = 0, W_1 = 0\}$, we have $X_1 = 0$ and $X_2 = W'$. Therefore

$$\begin{aligned}\mathsf{E}_{tu} f(X_1/t,\, X_2/t) &= \mathsf{E}_{tu} f(X_1/t,\, X_2/t)\, \mathbf{1}_{\{Z_1>0\}} + \mathsf{E}_{tu} f(X_1/t,\, X_2/t)\, \mathbf{1}_{\{Z_1=0, W_1>0\}} \\ &\quad + \mathsf{E}_{tu} f(X_1/t,\, X_2/t)\, \mathbf{1}_{\{Z_1=0, W_1=0\}} = A + B + C.\end{aligned}$$

For A, as $t \to \infty$, we have

$$\begin{aligned}A &= \mathsf{E} f\big(Z_1u + W_1/(t^2u),\, Z_2[Z_1u + W_1/(t^2u)] + W_2/[Z_1t^2u + W_1u^{-1}]\big)\, \mathbf{1}_{\{Z_1>0\}} \\ &\longrightarrow \mathsf{E} f(Z_1u,\, Z_1Z_2u)\, \mathbf{1}_{\{Z_1>0\}},\end{aligned}$$

while for B we obtain for $t \to \infty$,

$$B = \mathsf{E} f(W_1/t^2u,\, Z_2W_1/(t^2u) + W_2u/W_1)\, \mathbf{1}_{\{Z_1=0, W_1>0\}} \longrightarrow \mathsf{E} f(0,\, uW_2/W_1)\, \mathbf{1}_{\{Z_1=0, W_1>0\}}.$$

Finally for C,

$$C = \mathsf{E} f(0,\, W_2'/t)\, \mathbf{1}_{\{Z_1=0, W_1=0\}} = \mathsf{P}[Z_1 = 0, W_1 = 0]\, \mathsf{E} f(0,\, W_2'/t) \longrightarrow \mathsf{P}[Z_1 = 0, W_1 = 0]\, f(0,\, 0).$$

Observe that $\lim_{t\to\infty}[A + B + C] \neq \mathsf{E}_u f(T_1^*, T_2^*) = \mathsf{E} f(uZ_1, uZ_1Z_2)$. $\square$

In the final example, the conditional distributions of $t^{-1}\boldsymbol{X}$ converge to those of the tail chain $\boldsymbol{T}^*$, even though the regularity condition does not hold. This includes cases for which $G(\{0\}) = 0$ and $G(\{0\}) > 0$ with extremal boundary $y(t) \equiv 0$.

***Example* 6.3.** Let $\{(\xi_j, \eta_j),\, j \geq 1\}$ be iid copies of the non-degenerate random vector (ξ, η) on $[0,\infty)^2$. Taking $V = (\xi,\, \eta)$, consider a Markov chain which transitions according to the update function

$$\psi(y, V) = \xi(y + y^{-1})\, \mathbf{1}_{\{y>0\}} + \eta\, \mathbf{1}_{\{y=0\}} = \xi\, y + \big(\xi\, y^{-1}\, \mathbf{1}_{\{y>0\}} + \eta\, \mathbf{1}_{\{y=0\}}\big),$$

where the last expression is the canonical form. For $y > 0$, the transition kernel is

$$K\big(y\,,\,[0,x]\big) = \mathsf{P}\left[\xi(y + y^{-1}) \le x\right] = \mathsf{P}\left[\xi \le x/(y + y^{-1})\right].$$

For $t > 0$, $t^{-1}\psi(t,V) = \xi(1 + t^{-2}) \to \xi$ a.s., so $K \in D(G)$ with $G = \mathsf{P}[\xi \in \cdot]$. Note that $\phi(y,V) = \xi y^{-1}\,\mathbf{1}_{\{y>0\}} + \eta\,\mathbf{1}_{\{y=0\}}$ is unbounded near 0, implying that Segers' boundedness condition (4.2) does not hold. Also, our regularity condition (4.3) fails for K. To see this, write

$$K\big(tu_t\,,\,t(x,\infty)\big) = \mathsf{P}\left[\xi > x/(u_t + (t^2u_t)^{-1})\right].$$

Fix x so that $\mathsf{P}[\xi > x] > 0$ and choose $u_t = t^{-2}$. This yields $u_t + (t^2u_t)^{-1} = 1 + t^{-2}$, implying that

$$K\big(tu_t\,,\,t(x,\infty)\big) = \mathsf{P}\left[\xi > x/(1 + t^{-2})\right] \ge \mathsf{P}[\xi > x] > 0,$$

so (4.3) fails for K. However, since $K(t\,,\,\{0\}) = \mathsf{P}[\xi = 0] = G(\{0\})$, the choice $y(t) \equiv 0$ satisfies the definition of an extremal boundary (3.6), even if $G(\{0\}) > 0$. This leads to fdd convergence of $\mathsf{P}_{tu}[t^{-1}\boldsymbol{X} \in \cdot]$ to $\mathsf{P}_u[\boldsymbol{T}^* \in \cdot]$, and thus we learn that the conclusion (4.6) of Theorem 4.1 may hold without (4.3) being true.

We prove the fdd convergence for $m = 2$. For $u > 0$, and $X_0 = tu$, we have $X_1 = \xi_1(tu + (tu)^{-1})$ and $X_2 = \xi_2(X_1 + X_1^{-1})\,\mathbf{1}_{\{X_1>0\}} + \eta_2\,\mathbf{1}_{\{X_1=0\}}$. On $\{X_1 > 0\} = \{\xi_1 > 0\}$ we have $X_2 = \xi_2(X_1 + X_1^{-1})$. On $\{X_1 = 0\} = \{\xi_1 = 0\}$, we have $X_2 = \eta_2$. Thus, as $t \to \infty$,

$$\begin{aligned}\mathsf{E}_{tu} f(X_1/t,\, X_2/t)\,\mathbf{1}_{\{X_1>0\}} &= \mathsf{E}_{tu} f\big(\xi_1[u + (t^2u)^{-1}],\, \xi_2\xi_1[u + (t^2u)^{-2}] + \xi_2/(\xi_1[t^2u + 1/u])\big)\,\mathbf{1}_{\{X_1>0\}}\\ &\longrightarrow \mathsf{E} f(\xi_1 u,\, \xi_1\xi_2 u)\,\mathbf{1}_{\{\xi_1>0\}},\end{aligned}$$

while

$$\mathsf{E}_{tu} f(X_1/t,\, X_2/t)\,\mathbf{1}_{\{X_1=0\}} = \mathsf{E} f(0,\, \eta_2/t)\,\mathbf{1}_{\{\xi_1=0\}} \longrightarrow \mathsf{P}[\xi_1 = 0]\, f(0,0).$$

We conclude that

$$\mathsf{E}_{tu} f(X_1/t,\, X_2/t) \longrightarrow \mathsf{E} f(u\xi_1,\, u\xi_1\xi_2) = \mathsf{E}_u f(T_1^*,\, T_2^*). \qquad \square$$

7. Concluding Remarks

We have thus placed the traditional tail chain model for the extremes of a Markov chain in a more general context through the introduction of the boundary distribution H as well as the extremal boundary. A common application of the tail chain model is in deriving the weak limits of exceedance point processes for $\boldsymbol{X}$ [1, 18, 22]. We will shortly use our results to develop a detailed description of the clustering properties of extremes of Markov chains by means of such point processes. Furthermore, as we have not employed stationarity in our finite-dimensional results, we propose to substitute the inherent regenerative structure of a Harris recurrent Markov chain for the traditional assumption of stationarity. Also, it would be interesting to explore the implications of choices of H other than ϵ_0.

8. Appendix: Technical Lemmas

This section collects lemmas needed to prove convergence of integrals of the form $\int f_n\, d\mu_n$, assuming that $f_n \to f$ and $\mu_n \to \mu$ in their respective spaces. An example is the *second continuous mapping theorem* [2, Theorem 5.5, p. 34].

Lemma 8.1. *Assume $\mathbb{E}$ and $\mathbb{E}'$ are complete separable (cs) metric spaces, and for $n \ge 0$, $h_n : \mathbb{E} \to \mathbb{E}'$ are measurable. Put $A = \{x \in \mathbb{E} : h_n(x_n) \to h_0(x)$ whenever $x_n \to x\}$. If P_n, $n \ge 0$ are probability measures on $\mathbb{E}$ with $P_n \Rightarrow P_0$, and $h_n \to h_0$ almost uniformly in the sense that $P(A) = 1$, then $P_n \circ h_n^{-1} \Rightarrow P_0 \circ h_0^{-1}$ in $\mathbb{E}'$.*

The result provides a way to handle the convergence of a family of integrals.

Lemma 8.2. *In addition to the assumptions of Lemma 8.1, require $\mathbb{E}' = \mathbb{R}$ and $\{h_n,\, n \ge 0\}$ is uniformly bounded, so that $\sup_{n\ge0}\sup_{x\in\mathbb{E}} |h_n(x)| < \infty$.*

(a) We have

$$\int_{\mathbb{E}} h_n \, dP_n \longrightarrow \int_{\mathbb{E}} h_0 \, dP_0.$$

(b) Suppose additionally that $\mathbb{E}$ *is locally compact with a countable base (lccb), and* $\mu_n \xrightarrow{v} \mu_0$ *in* $\mathbb{M}_+(\mathbb{E})$ *with* $\mu_0(A^c) = 0$*. If there exists a compact set* $B \in \mathcal{K}(\mathbb{E})$ *with* $\mu_0(\partial B) = 0$ *such that* $h_n(x) = 0$, $n \geq 0$ *whenever* $x \notin B$ *(i.e.* B *is a common compact support of each* h_n*), then*

$$\int_{\mathbb{E}} h_n \, d\mu_n \longrightarrow \int_{\mathbb{E}} h_0 \, d\mu_0.$$

Proof. (a) If $X_n \sim P_n$ for $n \geq 0$, then $h_n(X_n) \Rightarrow h_0(X_0)$. The uniform boundedness of the h_n guarantees that $\mathsf{E}h_n(X_n) \to \mathsf{E}h_0(X_0)$.

(b) View B as a compact subspace of $\mathbb{E}$ inheriting the relative topology. Then, assuming $\mu(B) > 0$ to rule out a trivial case, define probabilities on B by $P_n(\cdot) = \mu_n(\cdot \cap B)/\mu_n(B)$, $n \geq 0$. Since $\mu_n(\cdot \cap B) \xrightarrow{v} \mu_0(\cdot \cap B)$ by Proposition 3.3 in [12], and B is compact, we get $P_n \Rightarrow P_0$. Denote by h_n', $n \geq 0$, the restriction of h_n to B. Observe that for any $x \in A \cap B$, we have $h_n'(x_n) \to h'(x)$ whenever $x_n \to x$ in B, and $P(A^c \cap B) \leq \mu(A^c)/\mu(B) = 0$. Therefore, apply part (a) to obtain

$$\int_{\mathbb{E}} h_n \, d\mu_n = \int_{\mathbb{E}} h_n \mathbf{1}_B \, d\mu_n = \mu_n(B) \int_B h_n' \, dP_n \longrightarrow \mu_0(B) \int_B h_0' \, dP_0 = \int_{\mathbb{E}} h_0 \, d\mu_0. \qquad \square$$

A convenient specialization of Lemma 8.2 (b) is the following.

Lemma 8.3. *Suppose* $\mathbb{E}$ *is lccb and* $\mu_n \xrightarrow{v} \mu$ *in* $\mathbb{M}_+(\mathbb{E})$*. If* $f : \mathbb{E} \to \mathbb{R}$ *is continuous and bounded, and* $B \in \mathbb{E}$ *is relatively compact with* $\mu(\partial B) = 0$*, then*

$$\int_B f \, d\mu_n \longrightarrow \int_B f \, d\mu.$$

Take $h_n = f\mathbf{1}_B$ for $n \geq 0$. Since $f\mathbf{1}_B$ is continuous except possibly on ∂B, we have $\mu(A^c) \leq \mu(\partial B) = 0$.

The next result is used to extend convergence of substochastic transition functions to multivariate regular variation on a larger space.

Lemma 8.4. *Let* $\mathbb{E} \subset [0, \infty]^m$ *and* $\mathbb{E}' \subset [0, \infty]^{m'}$ *be two nice (lccb) spaces. Suppose for* $t \geq 0$ *that* $\{p^{(t)}(\cdot\,, \cdot)\}_{t \geq 0}$*, are substochastic transition functions on* $\mathbb{E} \times \mathcal{B}(\mathbb{E}')$*. This means* $p^{(t)}(\cdot\,, B)$ *is a measurable function for any fixed* $B \in \mathcal{B}(\mathbb{E}')$*,* $p^{(t)}(x\,, \cdot)$ *is a measure for any* $x \in \mathbb{E}$*, and* $\sup_{t \geq 0} \sup_{u \in \mathbb{E}} p^{(t)}(u\,, \mathbb{E}') \leq 1$*. Assume there is a set* $A \subset \mathbb{E}$ *such that*

$$p^{(t)}\big(u_t\,, \cdot\big) \xrightarrow{v} p^{(0)}\big(u\,, \cdot\big) \qquad \text{in } \mathbb{M}_+(\mathbb{E}') \qquad (t \to \infty)$$

whenever $u_t \to u$ *in* $\mathbb{E}$ *and* $u \in A$*. Suppose also that* $\{\nu^{(t)}\}_{t \geq 0}$ *are measures on* $\mathbb{E}$ *such that* $\nu^{(0)}(A^c) = 0$*, and* $\nu^{(t)} \xrightarrow{v} \nu^{(0)}$ *in* $\mathbb{M}_+(\mathbb{E})$*. Then, defining measures* $\mu^{(t)}$ *for* $t \geq 0$ *on* $\mathbb{E} \times \mathbb{E}'$ *as*

$$\mu^{(t)}\big(du, dx\big) = \nu^{(t)}(du) p^{(t)}\big(u\,, dx\big)\,,$$

we have

$$\mu^{(t)} \xrightarrow{v} \mu^{(0)} \qquad \text{in } \mathbb{M}_+(\mathbb{E} \times \mathbb{E}') \qquad (t \to \infty).$$

Proof. Let $f \in \mathcal{C}_K^+(\mathbb{E} \times \mathbb{E}')$; without loss of generality assume f is supported on $K \times K'$, where $K \in \mathcal{K}(\mathbb{E})$ and $K' \in \mathcal{K}(\mathbb{E}')$. We have

$$\int_{\mathbb{E} \times \mathbb{E}'} \mu^{(t)}(du, dx)\; f(u, x) = \int_{\mathbb{E}} \nu^{(t)}(du) \int_{\mathbb{E}'} p^{(t)}\big(u\,, dx\big)\; f(u, x).$$

For $t \geq 0$, write

$$\varphi_t(u) = \int_{\mathbb{E}'} p^{(t)}\big(u\,,\,dx\big)\; f(u,x)$$

and suppose $u_t \to u_0$ with $u_0 \in A$; we verify that $\varphi_t(u_t) \to \varphi_0(u_0)$. Writing $g_t(x) = f(u_t, x)$, $t \geq 0$, we have $g_t(x_t) \to g_0(x_0)$ whenever $x_t \to x_0 \in \mathbb{E}'$ by the continuity of f. Also, the g_t are uniformly bounded by the bound on f and $g_t(x) = 0$ for all t whenever $x \notin K'$. Furthermore, without loss of generality we can assume that $p^{(0)}(u\,,\,\partial K') = 0$. Now apply Lemma 8.2 (b) to obtain

$$\varphi_t(u_t) = \int_{\mathbb{E}'} p^{(t)}\big(u_t\,,\,dx\big)\; g_t(x) \longrightarrow \int_{\mathbb{E}'} p^{(0)}\big(u\,,\,dx\big)\; g_0(x) = \varphi_0(u).$$

Since the $p^{(t)}$ are substochastic, and $\varphi_t(u) = 0$ for all t whenever $u \notin K$, the φ_t are uniformly bounded by the bound on f . Assume similarly that $\nu(\partial K) = 0$, and recall that $\nu(A^{\mathrm{c}}) = 0$. Apply Lemma 8.2 (b) once more to conclude as $t \to \infty$ that

$$\int_{\mathbb{E}\times\mathbb{E}'} \mu^{(t)}(du, dx)\; f(u,x) = \int_{\mathbb{E}} \nu^{(t)}(du)\; \varphi_t(u) \longrightarrow \int_{\mathbb{E}} \nu^{(0)}(du)\; \varphi_0(u) = \int_{\mathbb{E}\times\mathbb{E}'} \mu^{(0)}(du, x)\; f(u,x). \qquad \square$$

We conclude this section with a result used to verify the existence of the extremal boundary.

Lemma 8.5. *Suppose P_t, $t \geq 0$ are probability measures on a cs metric space $\mathbb{E}$ such that $P_t \Rightarrow P_0$, and let $A \subset \mathbb{E}$ be measurable. Then there exists a sequence of sets $A_t \downarrow \overline{A}$ such that $P_t(A_t) \to P_0(\overline{A})$.*

Remark. Note that if $P(\partial A) = 0$ then we can take $A_t = \overline{A}$. In the case of distribution functions $F_t \Rightarrow F$ on $\mathbb{R}^m$, taking $A = (-\infty, \boldsymbol{x}]$ and metric $\rho = \rho_\infty$ shows that for any $\boldsymbol{x} \in \mathbb{R}^m$ there exists $\boldsymbol{x}_t \downarrow \boldsymbol{x}$ such that $F_t(\boldsymbol{x}_t) \to F(\boldsymbol{x})$.

Proof. Let ρ be a metric on $\mathbb{E}$, and consider sets $A_\delta = \{x \,:\, \rho(x, A) \leq \delta\}$. Recall that $P_0(\partial A_\delta) = 0$ for all but a countable number of choices of δ, since $F(\delta) = P_0(A_\delta) - P_0(\overline{A})$ is a distribution function. First choose $\{\delta_k \,:\, k = 1, 2, \dots\}$ such that $0 < \delta_{k+1} \leq \delta_k \wedge 1/(k+1)$ and $P_0(\partial A_{\delta_k}) = 0$ for all k. Next, let $s_0 = 0$ and take $s_k \geq s_{k-1} + 1$, $k = 1, 2, \dots$ such that $P_t(A_{\delta_k}) > P_0(\overline{A}) - 1/k$ whenever $t \geq s_k$; this is possible since $P_t(A_{\delta_k}) \to P_0(A_{\delta_k}) \geq P_0(\overline{A})$ for all k. Finally, for $t > 0$ set

$$A(t) = A_{\delta_1}\, \mathbf{1}_{(0,s_1)}(t) + \sum_{k=1}^{\infty} A_{\delta_k}\, \mathbf{1}_{[s_k, s_{k+1})}(t).$$

We claim that $A(t) \downarrow \overline{A}$ and that $P_t(A(t)) \to P_0(\overline{A})$ as $t \to \infty$. It is clear that $A(t) \supset A(t')$ for $t \leq t'$, and $\cap_t A(t) = \cap_k A_{\delta_k} = \overline{A}$. On the one hand, for large t we have $A(t) \subset A_{\delta_k}$ for any k, so

$$\limsup_{t\to\infty}\; P_t(A(t)) \leq \limsup_{t\to\infty}\; P_t(A_{\delta_k}) \leq P_0(A_{\delta_k}).$$

Letting $k \to \infty$ shows that $\limsup_t P_t(A(t)) \leq P_0(\overline{A})$. On the other hand, if $k(t)$ denotes the value of k for which $s_k \leq t < s_{k+1}$, then

$$P_t(A(t)) = P_t(A_{\delta_{k(t)}}) > P_0(\overline{A}) - 1/k(t),$$

so $\liminf_t P_t(A(t)) \geq P_0(\overline{A})$. Combining these two inequalities shows that $P_t(A(t)) \to P_0(\overline{A})$. $\square$

References

[1] B. Basrak and J. Segers. Regularly varying multivariate time series. *Stochastic Processes and their Applications*, 119(4):1055–1080, 2009.

[2] P. Billingsley. *Convergence of probability measures.* Wiley, 1968.

[3] P. Billingsley. *Weak Convergence of Measures: Applications in Probability.* Society for Industrial and Applied Mathematics, Philadelphia, Pa., 1971. Conference Board of the Mathematical Sciences Regional Conference Series in Applied Mathematics, No. 5.

[4] P. Billingsley. *Convergence of probability measures.* Wiley-Interscience, 2nd edition, 1999.

[5] P. Bortot and S. Coles. A sufficiency property arising from the characterization of extremes of Markov chains. *Bernoulli*, 6(1):183–190, 2000.
[6] P. Bortot and S. Coles. Extremes of Markov chains with tail switching potential. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 65(4):851–867, 2003.
[7] P. Bortot and J.A. Tawn. Models for the extremes of Markov chains. *Biometrika*, 85(4):851, 1998.
[8] B. Das and S.I. Resnick. Conditioning on an extreme component: Model consistency with regular variation on cones. *Bernoulli*, 17(1):226–252, 2011.
[9] B. Das and S.I. Resnick. Detecting a conditional extreme value model. *Extremes*, 14:29–61, 2011.
[10] R.A. Davis and T. Hsing. Point process and partial sum convergence for weakly dependent random variables with infinite variance. *The Annals of Probability*, 23(2):879–917, 1995.
[11] L. de Haan and A. Ferreira. *Extreme value theory: an introduction*. Springer Verlag, 2006.
[12] P.D. Feigin, M.F. Kratz, and S.I. Resnick. Parameter estimation for moving averages with positive innovations. *The Annals of Applied Probability*, pages 1157–1190, 1996.
[13] J.E. Heffernan and S.I. Resnick. Limit laws for random vectors with an extreme component. *Annals of Applied Probability*, 17(2):537–571, 2007.
[14] J.E. Heffernan and J.A. Tawn. A conditional approach for multivariate extreme values. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 66(3):497–546, 2004.
[15] T. Hsing. Extreme value theory for multivariate stationary sequences. *Journal of Multivariate Analysis*, 29(2):274–291, 1989.
[16] O. Kallenberg. *Foundations of modern probability*. Springer Verlag, 1997.
[17] M.R. Leadbetter, G. Lindgren, and H. Rootzén. *Extremes and related properties of random sequences and processes*. Springer series in statistics. Springer-Verlag, 1983.
[18] R. Perfekt. Extremal behaviour of stationary Markov chains with applications. *The Annals of Applied Probability*, 4(2):529–548, 1994.
[19] R. Perfekt. Extreme value theory for a class of Markov chains with values in $\mathbb{R}^d$. *Advances in Applied Probability*, 29(1):138–164, 1997.
[20] S.I. Resnick. *Extreme values, regular variation, and point processes*. Springer Verlag, 2007.
[21] S.I. Resnick. *Heavy-tail phenomena: probabilistic and statistical modeling*. Springer Verlag, 2007.
[22] H. Rootzén. Maxima and exceedances of stationary Markov chains. *ADV. APPL. PROB.*, 20(2):371–390, 1988.
[23] J. Segers. Multivariate regular variation of heavy-tailed Markov chains. *Arxiv preprint math/0701411*, 2007.
[24] R.L. Smith. The extremal index for a Markov chain. *Journal of applied probability*, 29(1):37–45, 1992.
[25] R.L. Smith, J.A. Tawn, and S.G. Coles. Markov chain models for threshold exceedances. *Biometrika*, 84(2):249, 1997.
[26] S. Yun. The extremal index of a higher-order stationary Markov chain. *Annals of Applied Probability*, 8(2):408–437, 1998.
[27] S. Yun. The distributions of cluster functionals of extreme events in a dth-order Markov chain. *Journal of Applied Probability*, 37(1):29–44, 2000.

Sidney I. Resnick, School of ORIE, Rhodes Hall 284, Cornell University, Ithaca, NY 14853
E-mail address: sir1@cornell.edu

David Zeber, Department of Statistical Science, 301 Malott Hall, Cornell University, Ithaca NY 14853
E-mail address: dsz5@cornell.edu